GMM Estimation of Autoregressive Roots Near Unity with Panel Data

Hyungsik Roger Moon*
Department of Economics
University of Southern California
&

Peter C.B. Phillips
Cowles Foundation for Research in Economics
Yale University
University of Auckland and University of York

August, 2000

Abstract

This paper investigates a generalized method of moments (GMM) approach to the estimation of autoregressive roots near unity with panel data. The two moment conditions studied are obtained by constructing bias corrections to the score functions under OLS and GLS detrending, respectively. It is shown that the moment condition under GLS detrending corresponds to taking the projected score on the Bhattacharya basis, linking the approach to recent work on projected score methods for models with infinite numbers of nuisance parameters (Waterman and Lindsay, 1998). Assuming that the localizing parameter takes a nonpositve value, we establish consistency of the GMM estimator and find its limiting distribution. A notable new finding is that the GMM estimator has convergence rate $n^{1/6}$, slower than $\sqrt{n}$, when the true localizing parameter is zero (i.e., when there is a panel unit root) and the deterministic trends in the panel are linear. These results, which rely on boundary point asymptotics, point to the continued difficulty of distinguishing unit roots from local alternatives, even when there is an infinity of additional data.

JEL Classification: C22 & C23

Keywords and Phrases: Bias, boundary point asymptotics, GMM estimation, local to unity, moment conditions, nuisance parameters, panel data, pooled regression, projected score.

1 Introduction

Recent years have seen the introduction of several important panel data sets where the cross sectional dimension (say, n) and the time series dimension (say, T) are comparable

*The authors thank J. Owens for excellent research assistance, J. Hahn for helpful discussions, and R. Waterman for sharing with us his recent joint papers with B. Lindsay. Moon thanks the Academic Research Committee of UCSB for research support and Phillips thanks the NSF for research support under NSF Grant No. SBR 97-30295.

in magnitude. Some of these panel data sets, like the Penn World Tables, have time series components that are nonstationary. These features distinguish the new data from the characteristics that are conventionally assumed in the analysis of panel data.

Since the beginning of the 1990's, there has been ongoing theoretical and applied research on the use of large n and T panels allowing for nonstationarity in the data over time. The theoretical research includes the study of panel unit root tests (e.g., Quah, 1994, Levin and Lin, 1993, Im et al; 1996, Maddala and Wu, 1997, and Choi, 1999), panel cointegration tests (e.g., Pedroni, 1999, Binder et al), and the development of linear regression theories for panel estimators under nonstationarity (e.g., Pesaran and Smith, 1995, and Phillips and Moon, 1999). Applied research includes tests of growth convergence theories (Bernard and Jones, 1996), purchasing power parity relations (MacDonald, 1996, Oh, 1996, Pedroni, 1996, Wu, 1996, and Wu, 1997), and studies of the international links between savings and investment (Coakley et al, 1996 and Moon and Phillips, 1998).

Two recent papers by the authors (Moon and Phillips, 1999a & b) study panel regression models that allow for both deterministic trends and stochastic trends. When the deterministic trends in the nonstationary panel data are heterogeneous across individuals, Moon and Phillips (1999a) show that the maximum likelihood estimator (MLE) of the local to unity parameter in the stochastic trend is inconsistent. They call this phenomenon, which arises because of the presence of an infinite number of nuisance parameters, an incidental trend problem because it is analogous to the well-known incidental parameter problem in dynamic panels when T is fixed[1]. To solve the incidental trend problem, Moon and Phillips (1999b) propose various methods, including an iterative ordinary least squares (OLS) procedure and a double bias corrected estimator, and establish limit theories for these consistent estimators that can be used for statistical inference about the localizing parameter.

As a continuation of the two studies just mentioned, the present paper investigates a generalized method of moments (GMM) estimator of autoregressive roots near unity with panel data. We establish two moment conditions that form the basis for inference. The first moment condition is obtained by adjusting for the bias of the score function after conventional OLS detrending. The second moment condition is constructed by adjusting for the bias of the score function following GLS (or quasi-difference - QD) detrending. Interestingly, the second moment condition is shown to correspond to the Gaussian projected score, where the projection is taken on the so-called Bhattacharya basis that has been studied recently in the conventional incidental parameter problem by Waterman and Lindsay (1996, 1998) and Hahn (1998).

Consistency of the GMM estimator is proved under the assumption that the localizing parameter takes a nonpositive value. This condition is not too restrictive because most econometric models consider non-explosive autoregressive regression models. Nevertheless, the restriction does matter in deriving the limiting distribution of the estimator because it is possible that the true parameter lies on the boundary of the parameter set. The most interesting case is, of course, the pure unit root case where the true localizing parameter is zero. In this case, in establishing the limiting distribution we cannot use the conventional approach that approximates the first order condition because the true parameter could be on the boundary of the parameter set. To avoid this difficulty, we use the approach that takes a quadratic approximation of the nonlinear objective function and optimize it on the parameter set (c.f. Andrews, 1999, for some recent developments of estimation and inference in boundary problems).

One of the most interesting findings in the present paper is that the GMM estimator has slower convergence rate than $\sqrt{n}$ when the time series components in the panel have unit roots (i.e., the true localizing parameter is zero), and the deterministic trends are

[1] Lancaster (1998) provides a recent general survey of the incidental parameter problem in econometrics.

linear. In this case the convergence rate is actually $O(n^{1/6})$ rather than $O(\sqrt{n})$. This slow convergence rate arises because of lack of information in the moment conditions when there is a unit root, i.e., at the point $c = 0$ in the space of the localizing parameter. It points to the continued difficulty of distinguishing unit roots from local alternatives in the presence of deterministic trends even when there is an infinity of additional data from a cross section.

The paper is organized as follows. Section 2 lays out the model and gives the basic assumptions that are maintained thought the paper. In section 3 we introduce two moment conditions and prove that the second moment condition corresponds to a Gaussian projected score on the Bhattacharya basis. In Section 4 we establish consistency of the GMM estimator and obtain the limiting distributions of the GMM estimator when the true parameter is less than zero and equal to zero. The appendix contains technical derivations and proofs of the results in the main text.

2 Model and Assumptions

The model considered here is the panel system written in components form

$$\begin{aligned} z_{it} &= \beta_i' g_{pt} + y_{it} \\ y_{it} &= \rho y_{it-1} + \varepsilon_{it}, \end{aligned} \tag{1}$$

where the autoregressive coefficient

$$\rho = \exp\left(\frac{c}{T}\right) \sim 1 + \frac{c}{T},$$

is local to unity and the deterministic trend

$$g_{pt} = (t, \ldots, t^p)' : \ (p \times 1) \ \text{polynomial trend vector.}$$

Let β_{i0} and $\rho_0 = 1 + \frac{c_0}{T}$ denote the true parameters. The main interest of the paper is to find a consistent estimation procedure for the localizing parameter c_0. A case of special interest is the panel unit root model where $c_0 = 0$.

In practice, the most widely used trend in empirical applications is the linear trend, when $g_{1t} = t$ in (1). In later sections of the paper as part of the asymptotic development we need to verify some properties of complicated nonlinear functions of c that depend on the trend g_{pt}. These functions are so complicated that it is very difficult to establish general analytic results under the set up of the general polynomial trend function $g_{pt} = (t, \ldots, t^p)'$. Instead, we rely on numerical methods for this part of the analysis. And to assist the analytic development, we restrict our attention to the following two cases: (i) $g_{1t} = t$ and (ii) $g_{2t} = \left(t, t^2\right)'$. The set up is formalized as follows:

Assumption 1 (Trend Formulation)
The polynomial trend in model (1) is either (i) $g_{1t} = t$ or (ii) $g_{2t} = \left(t, t^2\right)'$.

Assumption 2 (Error Condition) ε_{it} are linear processes satisfying the following conditions.

(a) $\varepsilon_{it} = \sum_{j=0}^{\infty} C_{ij} u_{it-j}$, where u_{it} are iid across i over t with $Eu_{it} = 0$, $Eu_{it}^2 = 1$, and $Eu_{it}^4 = \sigma_{u,4} < \infty$.

(b) C_{ij} are sequence of real numbers with $\bar{C}_j = \sup_i |C_{ij}| < \infty$ and $\sum_{j=0}^{\infty} j^b \bar{C}_j < \infty$ for some $b > 2$.

Assumption 3 (Initial Condition)

(a) $y_{i0} = z_{i0}$ for all i

(b) $E\sup_i |y_{i0}|^{\kappa} < \infty$ for some $\kappa > 4$.

Assumption 4 (Parameter Set)

(a) The localizing parameter c takes a value in a compact subset $C = [\bar{c}, 0] \subset \mathbb{R}$, where $\bar{c} < 0$.

(b) The true localizing parameter c_0 is in the set $C_0 = (\bar{c}, 0]$.

Assumption 4(a) restricts the parameter set $C = [\bar{c}, 0]$ to be non-positive. This restriction is made because in most econometrics application, $|\rho| < 1$ or $\rho = 0$ is of most interest. When the true parameter $c_0 = 0$, the model becomes nonstandard in the sense that the true parameter is on the boundary of the parameter set. Section 5 explores the implications of the boundary point aspect of this case.

Let $C_i = \sum_{j=0}^{\infty} C_{ij}$, $\Omega_i = C_i^2$, and $\Lambda_i = \sum_{j=1}^{\infty} C_{i0}C_{ij}$. Ω_i and Λ_i are the long-run variance and the one-sided covariance of the error process ε_{it}, respectively. The next assumption is about the limits of the averages of the individual long-run variances and covariances.

Assumption 5 (Long Run Variances)

(a) $\inf_i \Omega_i > 0$

(b) $\Omega = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \Omega_i$ is finite.

(c) $\Psi^2 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \Omega_i^2$ is finite.

(d) $\Lambda = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \Lambda_i$ is finite.

In most applications, the long-run variances Ω_i and Λ_i are not known and consistent estimates of Ω_i and Λ_i are required. A widely used method is to employ a kernel estimation approach (c.f., Park and Phillips, 1988). Once we obtain consistent estimates of Ω_i and Λ_i, we can average them to produce consistent estimates of the quantities Λ and Ω. Specifically, suppose that $\hat{u}_{it}$ is a regression residual of model (1) or model (4). Define the sample covariances $\hat{\Gamma}_i(j) = \frac{1}{T}\sum \hat{u}_{it}\hat{u}_{it+j}$, where the summation is defined over $1 \le t, t+j \le T$. Then, the kernel estimators for $\hat{\Lambda}_i$ and $\hat{\Omega}_i$ are:

$$\hat{\Lambda}_i = \sum_{j=1} w\left(\frac{j}{K}\right)\hat{\Gamma}_i(j), \tag{2}$$

$$\hat{\Omega}_i = \sum_{j=-T} w\left(\frac{j}{K}\right)\hat{\Gamma}_i(j), \tag{3}$$

where $w(\cdot)$ is a kernel function with $w(0) = 1$ and K is a lag truncation parameter. Truncation occurs when $w\left(\frac{j}{K}\right) = 0$ for $|j| \ge K$. Averaging over cross section observations now leads to consistent estimators of Λ and Ω, viz.,

$$\hat{\Lambda} = \frac{1}{n}\sum_{i=1} \hat{\Lambda}_i \text{ and } \hat{\Omega} = \frac{1}{n}\sum_{i=1} \hat{\Omega}_i.$$

We assume that the estimates $\hat{\Lambda}_i$ and $\hat{\Omega}_i$ have the following desirable properties. Examples of such estimates $\hat{\Lambda}_i$ and $\hat{\Omega}_i$ are found in Moon and Phillips (1999b), and we will not pursue this aspect of the theory further here.

Assumption 6 (Long Run Variance Estimation) Assume[2] that as $(n, T \to \infty)$ with $\frac{n}{T} \to 0$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| \hat{\Lambda}_i - \Lambda_i \right| \text{ and } \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| \hat{\Omega}_i - \Omega_i \right| = o_p(1).$$

3 Moment Conditions

This section develops two moment conditions that will be used in GMM estimation of c_0. The central idea is to correct for the biases in the OLS detrended regression and in GLS detrended regression, a process that leads to two different moment conditions. It turns out that the second moment condition is equivalent to a particular form of projected score in the Gaussian version of model (1). The projection is on the Bhattarcharya basis (Bhattacharyya, 1946 and Waterman and Lindsay, 1996) and this correspondence is explored in the final part of this section.

3.1 The First Moment Condition

We start by writing Model (1) in augmented regression format as

$$z_{it} = \rho_0 z_{it-1} + \delta_{i0} + \gamma_{i0}' \tilde{g}_{pt} + \varepsilon_{it}, \tag{4}$$

where

$$\begin{aligned}
\delta_{i0} &= \rho_0 \bar{\beta}_{i0}' \iota_p, \\
\gamma_{i0} &= \bar{\beta}_{i0}' \Upsilon_T(c_0), \\
\iota_p &= \left(-1, (-1)^2, \dots, (-1)^p\right)', \\
\Upsilon_T(c_0) &= (p \times p) \text{ matrix dedending on } c_0 \text{ and } T.
\end{aligned}$$

The augmented format (4) has the drawback that linear regression leads to inefficient trend elimination, but it has the advantage that the detrended data is invariant to the trend parameters in (1). The first moment condition uses the augmented formation (4) and the second moment condition uses model (1).

The following notation is defined to assist with the analysis of the trend function asymptotics and it will be used subsequently throughout the paper. Let

$$\begin{aligned}
\underline{\nu}_{i0} &= \left(\delta_{i0}, \gamma_{i0}'\right)', \\
\tilde{g}_{pt} &= \left[1, g_{pt}'\right]', \qquad g_p(r) = (r, \dots, r^p)', \qquad \tilde{g}_p(r) = \left[1, g_p(r)'\right]', \\
G_{pT} &= \left[g_{p1}', \dots, g_{pT}'\right]', \qquad \tilde{G}_{pT,-1} = \left[\tilde{g}_{p0}', \dots, \tilde{g}_{pT-1}'\right]', \qquad \tilde{G}_{pT} = \left[\tilde{g}_{p1}', \dots, \tilde{g}_{pT}'\right]', \\
\tilde{M}_{pT} &= I_T - \tilde{G}_{pT} \left(\tilde{G}_{pT}' \tilde{G}_{pT}\right)^{-1} \tilde{G}_{pT}', \\
D_{pT} &= \operatorname{diag}(T, \dots, T^p), \qquad \tilde{D}_{pT} = \operatorname{diag}(1, D_T), \\
h_{pT}(t, s) &= D_{pT}^{-1} g_{pt}' \left(\frac{1}{T} \sum_{t=1}^{T} D_{pT}^{-1} g_{pt} g_{pt}' D_{pT}^{-1} \right)^{-1} g_{ps} D_{pT}^{-1},
\end{aligned}$$

[2] Usually, the lag truncation parameter K in (2) and 3 tends to infinity as n, T increase to infinity together, under a certain regularity condition. For example, Moon and Phillips (1999b) impose the condition that $\frac{nK}{T} \to 0$ as $(n, T \to \infty)$ with $\frac{n}{T} \to 0$. This regularity condition is required for the asymptotics underlying Assumption 6.

$$\bar{h}_{pT}(t;s) = D_{pT}^{-1}\bar{g}_{pt}'\left(\frac{1}{T}\sum_{t=1} D_{pT}^{-1}\bar{g}_{pt}\bar{g}_{pt}'D_{pT}^{-1}\right)^{-1}\bar{g}_{ps}D_{pT}^{-1};$$

$$h_p(r;s) = g_p'(r)\left(\int_0^1 g_p(r)\,g_p(r)'dr\right)^{-1} g_p(s);$$

$$\bar{h}_p(r;s) = \bar{g}_p'(r)\left(\int_0^1 \bar{g}_p(r)\,\bar{g}_p(r)'dr\right)^{-1} \bar{g}_p(s):$$

Write $z_i = (z_{i1},\ldots,z_{iT})'$; $z_{i,-1} = (z_{i0},\ldots,z_{iT-1})'$; and $\varepsilon_i = (\varepsilon_{i1},\ldots,\varepsilon_{iT})'$: Let

$$\underline{z}_i = \bar{M}_{pT} z_i;\ \underline{\varepsilon}_i = \bar{M}_{pT}\varepsilon_i;\ \underline{z}_{i,-1} = \bar{M}_{pT} z_{i,-1}:$$

Then, it is straightforward to show that

$$\underline{z}_i = \underline{y}_i \text{ and } \underline{z}_{i,-1} = \underline{y}_{i,-1};$$

where

$$\underline{y}_i = \bar{M}_{pT} y_i;\ \underline{y}_{i,-1} = \bar{M}_{pT} y_{i,-1};$$

$y_i = (y_1,\ldots,y_T)'$; and $y_{i,-1} = (y_0,\ldots,y_{T-1})'$: For $t \geq 2$ we let

$$\left(\underline{z}_{i,-1}\right)_t = z_{it-1} - \frac{1}{T}\sum_{s=1}\bar{h}_{pT}(t;s)\, z_{is-1}$$

be the t^{th} element of $\underline{z}_{i,-1}$; and assume $\left(\underline{z}_{i,-1}\right)_1 = z_{i0} = y_{i0}$:

One straightforward procedure of estimating c_0 (equivalently ρ_0) is to first eliminate the unknown trends $\delta_{i0} + \gamma_{i0}' g_t$ by taking OLS regression residuals and then apply pooled least squares with an appropriate bias correction for the serial correlation of ε_{it}, calling this method iterative OLS. However, as noted by Moon and Phillips (1999b), this iterative OLS procedure yields inconsistent estimation of c_0 due to a nondegenerating asymptotic bias between the detrended regressor and the detrended error term.

The first moment condition is obtained simply by subtraction of this asymptotic bias term in an iterative OLS procedure. More specifically, we write Model (4) in vector notation as

$$z_i = \rho_0 z_{i,-1} + G_{pT}\nu_{i0} + \varepsilon_i:$$

Multiplying $\bar{M}_{pT}$ to the both sides of the equation, we have

$$\underline{z}_i = \rho_0\,\underline{z}_{i,-1} + \underline{\varepsilon}_i;$$

where $\underline{z}_i; \underline{z}_{i,-1}$; and $\underline{\varepsilon}_i$ are OLS detrended versions of z_i; $z_{i,-1}$; and ε_i; respectively. In general, the detrended regressor vector $\underline{z}_{i,-1}$ and the detrended error vector $\underline{\varepsilon}_i$ are correlated.

The first moment condition is found by correcting for the bias due to the correlation between $\underline{z}_{i,-1}$ and $\underline{\varepsilon}_i$: We will use $m_{1,iT}(c)$ to denote the data moment that appears in the first moment condition. It is defined as follows:

$$m_{1;iT}(c) = \frac{1}{T}\left(\underline{z}_i - \left(1+\frac{c}{T}\right)\underline{z}_{i,-1}\right)'\underline{z}_{i,-1} - \hat{\sigma}_i\,\omega_{1T}(c) - \hat{\lambda}_i \tag{5}$$

$$
\begin{aligned}
&= \frac{1}{T}\underline{\varepsilon}_i'\underline{y}_{i,-1} - (c - c_0)\frac{1}{T^2}\underline{y}_{i,-1}'\underline{y}_{i,-1} - \hat{\beta}_i\,\omega_{1T}(c) - \hat{\Lambda}_i \\
&= \frac{1}{T}\sum_{t=1}\varepsilon_{it}y_{i,t-1} - \frac{1}{T^2}\sum_{t=1}\sum_{s=1}\varepsilon_{it}y_{i,s-1}h_{pT}(t,s) - (c - c_0)\frac{1}{T^2}\sum_{t=1}\left(\underline{y}_{i,-1}\right)_t^2 \\
&\quad - \hat{\beta}_i\,\omega_{1T}(c) - \hat{\Lambda}_i,
\end{aligned}
$$

where

$$
\omega_{1T}(c) = -\frac{1}{T^2}\sum_{t=2}\sum_{s=1}^{t-1} e^{\left(\frac{t-s-1}{T}\right)c}h_{pT}(t,s),
$$

and $\underline{\varepsilon}_{it}$ and $\left(\underline{y}_{i,-1}\right)_t$ are the t^{th} elements of $\underline{\varepsilon}_i$ and $\underline{y}_{i,-1}$, respectively. The terms $\hat{\beta}_i\,\omega_{1T}(c)$ and $\hat{\Lambda}_i$ correct for the asymptotic bias that arises from the correlation between $\underline{\varepsilon}_{it}$ and $\left(\underline{y}_{i,-1}\right)_t$.

Since the bias correction terms $\hat{\beta}_i\,\omega_{1T}(c)$ and $\hat{\Lambda}_i$ are approximations of the mean of $\frac{1}{T}\underline{\varepsilon}_i'\underline{y}_{i,-1}$, $E(m_{1,iT}(c_0))$ is not exactly zero but it is asymptotically zero, in general. However, $m_{1,iT}(c)$ has a simple limiting form that delivers an exact moment condition. When T is large, it is easy to find that the distribution of $m_{1,iT}(c)$ is close to that of

$$
\beta_i\left(\int_0^1 \underline{J}_{c_0,i}(r)\,dW_i(r) - (c - c_0)\int_0^1 \underline{J}_{c_0,i}(r)^2\,dr - \omega_1(c)\right),
$$

where $J_{c_0,i}(r) = \int_0^r e^{c_0(r-s)}dW_i(s)$ is a diffusion, $W_i(r)$ is standard Brownian Motion, $\underline{J}_{c_0,i}(r) = J_{c_0,i}(r) - \int_0^1 J_{c_0,i}(s)h_p(r,s)\,ds$, and $\omega_1(c) = -\int_0^1\int_0^r e^{c_0(r-s)}h_p(r,s)\,ds\,dr$. Since

$$
E\left(\int_0^1 \underline{J}_{c_0,i}(r)\,dW_i(r)\right) = \omega_1(c_0),
$$

it follows that when $c = c_0$

$$
E\left(\beta_i\left(\int_0^1 \underline{J}_{c_0,i}(r)\,dW_i(r) - (c - c_0)\int_0^1 \underline{J}_{c_0,i}(r)^2\,dr - \omega_1(c)\right)\right) = 0,
$$

giving the moment condition directly for this limiting form of $m_{1,iT}(c_0)$.

3.2 The Second Moment Condition

Before we discuss the second moment condition, we introduce the following notation. Let

$$
\begin{aligned}
\Delta_c &= 1 - \left(1 + \frac{c}{T}\right)L, \text{ where } L \text{ is the lag operator,} \\
F_{pT} &= \operatorname{diag}\left[1, T, \ldots, T^{p-1}\right] = \frac{1}{T}D_{pT}, \qquad \Delta_c g_{pt} = F_{pT}^{-1}\Delta_c g_{pt} \\
\dot{g}_p(r) &= \frac{d}{dr}g_p(r) = \left[1, 2r, \ldots, pr^{p-1}\right]', \qquad \dot{g}_{pc}(r) = \dot{g}_p(r) - cg_p(r), \\
A_{pT}(c) &= \frac{1}{T}\sum_{t=1}\Delta_c g_{pt}\,\Delta_c g_{pt}', \qquad A_p(c) = \int_0^1 \dot{g}_{pc}(r)\,\dot{g}_{pc}(r)'\,dr, \\
B_{pT}(c) &= \frac{1}{T}\sum_{t=1}\Delta_c g_{pt}\,g_{pt-1}'D_{pT}^{-1}, \qquad B_p(c) = \int_0^1 \dot{g}_{pc}(r)\,g_p(r)'\,dr.
\end{aligned}
$$

The second moment condition is obtained from the efficiently detrended regression equation. According to Canjels and Watson (1997) and Phillips and Lee (1996), the trend coefficient in the model (1) can be efficiently estimated in the time domain by employing a GLS procedure that amounts to quasi-differencing the data with the operator Δ_c. That is, when the localizing parameter c is known, the asymptotically efficient estimator of β_i in (1) is

$$\hat{\beta}_i(c) = \left(\sum_{t=1}^{T} \Delta_c g_{pt} \Delta_c g_{pt}'\right)^{-1} \left(\sum_{t=1}^{T} \Delta_c g_{pt} \Delta_c z_{it}\right):$$

Denoting $y_{it}(\beta_i) = z_{it} - \beta_i' g_{pt}$, we now write

$$\hat{\beta}_i(c) = \beta_{i0} + \left(\sum_{t=1}^{T} \Delta_c g_{pt} \Delta_c g_{pt}'\right)^{-1} \left(\sum_{t=1}^{T} \Delta_c g_{pt} y_{it}(\beta_{i0})\right):$$

Define $\varepsilon_{it}(c; \beta_{i0}) = \Delta_c z_{it} - \beta_i' \Delta_c g_{pt}$.

The second moment function $m_{2;iT}(c)$ is defined as

$$m_{2;iT}(c) = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it}\left(c; \hat{\beta}_i(c)\right) y_{it-1}\left(\hat{\beta}_i(c)\right) - \hat{\lambda}_{i,T}(c) - \hat{\Omega}_i; \qquad (6)$$

where

$$\lambda_T(c) = -\operatorname{tr}\left(A_{pT}(c)^{-1} \frac{1}{T^2} \sum_{t=2}^{T} \sum_{s=1}^{t-1} e^{\left(\frac{t-s-1}{T}\right)c} \Delta_c g_{pt} \Delta_c g_{ps}'\right):$$

Notice that $y_{it-1}\left(\hat{\beta}_i(c)\right)$ is the GLS regression residual of the regression equation $z_{it} = \beta_i' g_t + y_{it}$ and $\varepsilon_{it}\left(c; \hat{\beta}_i(c)\right)$ is the OLS regression residual of the quasi-differenced equation $\Delta_c z_{it} = \beta_i' \Delta_c g_{pt} + \Delta_c y_{it}$. In the second moment function $m_{2;iT}(c)$ we correct for the asymptotic bias of $\frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}\left(c; \hat{\beta}_i(c)\right) y_{it-1}\left(\hat{\beta}_i(c)\right)$ by substracting off the estimates $\hat{\lambda}_{i,T}(c)$ and $\hat{\Omega}_i$.

Recently, Moon and Phillips (1999a) showed that the Gaussian MLE of the panel regression model (2) with linear incidental trends is inconsistent. The main reason for inconsistency of the MLE is that the concentrated score of the (standardized) Gaussian likelihood function, $\frac{1}{n}\sum_{i=1}^{n} \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}\left(c; \hat{\beta}_i(c)\right) y_{it-1}\left(\hat{\beta}_i(c)\right)$, has non-zero mean in the limit. In the second moment formulation of $m_{2;iT}(c)$, by subtracting off the estimates $\hat{\lambda}_{i,T}(c)$ and $\hat{\Omega}_i$, we eliminate the asymptotic bias of the concentrated Gaussian score function.

3.3 The Relationship between the Second Moment Condition and the Projected Score

This section shows that the second moment function $m_{2;iT}(c)$ is a projected score of the panel regression model (1) with Gaussian errors. Suppose that the error process ε_{it} in the model (1) is an iid standard normal process across i and over t. For convenience we assume that $z_{i0} = y_{i0} = 0$ for all i.

Under general regularity conditions, it is well known that the asymptotic properties of the MLE, and most notably its consistency, are closely related to the unbiasedness of the score function at the true parameter. However, it is also well known that in dynamic panel regression models with incidental parameters the MLE is not consistent (e.g., see Neyman and Scott, 1948, and Nickel, 1981) as $n \to \infty$ with T fixed. Recently, Moon and Phillips (1999b) found that this incidental parameter problem also arises in the nonstationary panel regression models with incidental trends when both $n \to \infty$ and $T \to \infty$, to wit in models such as (1).

The main reason for the inconsistency of the MLE is that the score function in an incidental trend model has a bias at the true parameter. Therefore, in order to obtain a consistent estimate, one needs to correct for the bias in the score function. One recently investigated method to correct for this bias is to use a projected score function, where the projection is taken onto the so-called Bhattacharyya basis. The resulting approach is called "a projected score method".

To define a projected score in the present case, we introduce the following notation. Let

$$f_i(z_i;c;\beta_i) = \left(\frac{1}{\sqrt{2\pi}}\right)^T \exp\left(-\frac{1}{2}\sum_{t=1}^T \left[\Delta_c z_{it} - \beta_i' \Delta_c g_{pt}\right]^2\right) \tag{7}$$

: the joint density of z_i,

$$U_{1i} = \frac{\partial f_i/\partial c}{f_i}; \qquad V_{1i} = \frac{\partial f_i/\partial \beta_i}{f_i};$$

$$V_{2i} = \frac{\frac{\partial^2 f_i}{\partial \beta_i \partial \beta_i'}}{f_i} + \frac{\partial f_i}{\partial \beta_i}\frac{\partial f_i}{\partial \beta_i'}; \qquad V_i = \begin{pmatrix} V_{1i} \\ D_p^+ \operatorname{vec} V_{2i} \end{pmatrix};$$

where $D_p^+ = \left[D_p' D_p\right]^{-1} D_p'$ and D_p is the duplication matrix. In the statistics literature, V_{1i} and V_{2i} are known as the Bhattacharyya basis of order 1 and 2, respectively (e.g., Bhattacharyya, 1946 and Waterman and Lindsay, 1996). The projected score U_{2i} is defined as the residual in the L_2- projection of U_{1i} on the closed linear space spanned by V_{1i} and V_{2i}; i.e.,

$$U_{2i} = U_{1i} - \xi_1' V_{1i} - \xi_2' D_p^+ (\operatorname{vec} V_{2i}). \tag{8}$$

Recently, using the projected score method, Waterman and Lindsay (1998) and Hahn (1998) were able to solve similar nuisance parameter problems in the classical Neyman and Scott panel regression model and in a simple dynamic panel regression model with fixed effects, respectively.

When the joint density of z_i is given in (7); U_{1i}; V_{1i}; and V_{2i} are found to be

$$U_{1i}(c;\beta_i) = \frac{1}{T}\sum_{t=1}^T \varepsilon_{i,t}(c;\beta_i)\, y_{i,t-1}(\beta_i);$$

$$V_{1i}(c;\beta_i) = \sum_{t=1}^T \varepsilon_{i,t}(c;\beta_i)\, \Delta_c g_{pt};$$

$$V_{2i}(c;\beta_i) = -\sum_{t=1}^T \Delta_c g_{pt} \Delta_c g_{pt}' + \left(\sum_{t=1}^T \varepsilon_{i,t}(c;\beta_i)\, \Delta_c g_{pt}\right)\left(\sum_{t=1}^T \varepsilon_{i,t}(c;\beta_i)\, \Delta_c g_{pt}\right)'.$$

After some algebra, we obtain

$$E\left(V_{1i} \otimes \operatorname{vec} V_{2i}\right) = 0$$

and

$$E V_{1i} U_{1i} = 0.$$

So, the two L_2-projection coefficients λ_1 and λ_2 in (8) are given by

$$\lambda_1 = [EV_{1i}V_{1i}']^{-1} EV_{1i}U_{1i} = 0;$$

and

$$\lambda_2 = \left[D_p^+ E(\mathrm{vec}V_{2i})(\mathrm{vec}V_{2i})' D_p^{+\prime} \right]^{-1} D_p^+ E(\mathrm{vec}V_{2i}) U_{1i}.$$

Also, after some lengthy calculation, we find that

$$\begin{aligned} & E(\mathrm{vec}V_{2i})(\mathrm{vec}V_{2i})' \\ = & \sum_{t=1}^{T}\sum_{s=1}^{T} \left(\Delta_c g_{pt} \Delta_c g_{pt}' \otimes \Delta_c g_{ps} \Delta_c g_{ps}' \right) + \sum_{t=1}^{T}\sum_{s=1}^{T} \left(\Delta_c g_{pt} \Delta_c g_{ps}' \otimes \Delta_c g_{ps} \Delta_c g_{pt}' \right); \end{aligned}$$

and

$$\begin{aligned} & E(\mathrm{vec}V_{2i}) U_{1i} \\ = & \frac{1}{T} \sum_{t=2}^{T}\sum_{s=1}^{t-1} \left[\Delta_c g_{pt} \otimes \Delta_c g_{ps} + \Delta_c g_{ps} \otimes \Delta_c g_{pt} \right] e^{\left(\frac{t-s-1}{T}\right)c}. \end{aligned}$$

Therefore, the projected score $U_{2i}(c;\beta_i)$ is

$$\begin{aligned} & U_{2i}(c;\beta_i) \\ = & \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{i,t}(\beta_i;c) y_{i,t-1}(\beta_i) + \lambda_2' D_p^+ \sum_{t=1}^{T} (\Delta_c g_{pt} \otimes \Delta_c g_{pt}) \\ & - \lambda_2' D_p^+ \left(\sum_{t=1}^{T} \varepsilon_{i,t}(c;\beta_i) \Delta_c g_{pt} \right) \otimes \left(\sum_{s=1}^{T} \varepsilon_{i,s}(c;\beta_i) \Delta_c g_{ps} \right); \end{aligned}$$

where

$$\begin{aligned} & \lambda_2 \\ = & \left[\sum_{t=1}^{T}\sum_{s=1}^{T} D_p^+ \left\{ \left(\Delta_c g_{pt} \Delta_c g_{pt}' \otimes \Delta_c g_{ps} \Delta_c g_{ps}' \right) + \left(\Delta_c g_{pt} \Delta_c g_{ps}' \otimes \Delta_c g_{ps} \Delta_c g_{pt}' \right) \right\} \left(D_p^+ \right)' \right]^{-1} \\ & \times \frac{1}{T} \sum_{t=2}^{T}\sum_{s=1}^{t-1} D_p^+ \left[\Delta_c g_{pt} \otimes \Delta_c g_{ps} + \Delta_c g_{ps} \otimes \Delta_c g_{pt} \right] e^{\left(\frac{t-s-1}{T}\right)c}. \end{aligned}$$

Since β_i in U_{2i} is unknown, we replace it with the estimate

$$\hat{\beta}_i(c) = \left(\sum_{t=1}^{T} \Delta_c g_{pt} \Delta_c g_{pt}' \right)^{-1} \left(\sum_{t=1}^{T} \Delta_c g_{pt} \Delta_c z_{it} \right).$$

Then, we have the following concentrated projected score

$$U_{2i}\left(c; \hat{\beta}_i(c)\right) = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{i,t}\left(\hat{\beta}_i(c); c\right) y_{i,t-1}\left(\hat{\beta}_i(c)\right) + \lambda_2' D_p^+ \sum_{t=1}^{T} (\Delta_c g_{pt} \otimes \Delta_c g_{pt}); \quad (9)$$

because $\sum_{t=1}^{T} \varepsilon_{i,t}\left(c; \hat{\beta}_i(c)\right) \Delta_c g_{pt} = 0$.

Now, when the error process ε_{it} is iid$(0,1)$ across i and over t, the second moment function $m_{2;iT}(c)$ is

$$m_{2;iT}(c) = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}\left(c;\hat{\theta}_i(c)\right)\left(y_{it-1}\left(\hat{\theta}_i(c)\right) - \lambda_T(c)\right).$$

The following lemma states that the bias correction term $-\lambda_T(c)$ in $m_{2;iT}(c)$ is equivalent to $\xi_2' D_p^+ \sum_{t=1}^{T}\left(\Delta_c g_{pt} - \overline{\Delta_c g_{pt}}\right)$. Thus, we conclude that the second moment function actually corresponds to the concentrated projected score function of the Gaussian model.

Lemma 1 (Equivalence) Suppose that the errors in model 1 are iid normal with mean zero and variance 1 across i and over t and $y_{i0} = z_{i0} = 0$ for all i. Then, the second moment condition $m_{2;iT}(c)$ is equivalent to the concentrated projected score function $U_{2i}\left(c;\hat{\theta}_i(c)\right)$.

4 GMM Estimation and Asymptotics

This section investigates the asymptotic properties of a GMM estimator of c that is based on the two moment conditions introduced in the previous section. Let

$$M_{nT}(c) = \frac{1}{n}\sum_{i=1}^{n} m_{iT}(c),$$

where

$$m_{iT}(c) = \begin{pmatrix} m_{1;iT}(c) \\ m_{2;iT}(c) \end{pmatrix},$$

and where $m_{1;iT}(c)$ and $m_{2;iT}(c)$ are defined in (5) and (6), respectively. Let $\hat{W}$ be a (2×2) random weight matrix and B_{nT} be a sequence of real numbers that converges to infinity as $(n,T \to \infty)$. The GMM estimator $\hat{c}$ for the unknown parameter c_0 in (1) is defined as the extremum estimator for which

$$Z_{nT}(\hat{c}) \leq \min_{c \in C} Z_{nT}(c) + o_p\left(B_{nT}^{-2}\right), \tag{10}$$

where

$$Z_{nT}(c) = M_{nT}(c)' \hat{W} M_{nT}(c).$$

Since the objective function $Z_{nT}(c)$ is continuous in c and the parameter set C assumed to be compact, it is possible to find a global minimum of $Z_{nT}(c)$ over the parameter set C. The main purpose in allowing for an $o_p\left(B_{nT}^{-1}\right)$ deviation bound from the global minimum $\min_{c \in C} Z_{nT}(c)$ is to reduce the computational burden and allow for potential numerical computational errors within a range of $o_p\left(B_{nT}^{-1}\right)$. Later in this paper, depending on the convergence order of $\hat{c}$ to c_0, we will determine the sequence B_{nT}.

4.1 Consistency of the GMM Estimator

Define

$$M(c) = \begin{pmatrix} m_1(c) \\ m_2(c) \end{pmatrix},$$

where

$$m_1(c) = \omega_1(c_0) - \omega_1(c) - (c - c_0)\,\omega_2(c_0);$$

$$
\begin{aligned}
\omega_1(c) &= -\int_0^1\int_0^r e^{c(r-s)}\bar{h}_p(r,s)\,ds\,dr;\\
\omega_2(c_0) &= -\frac{1}{2c_0}\left(1+\frac{1}{2c_0}\left[1-e^{2c_0}\right]\right)\\
&\quad -\int_0^1\int_0^1 e^{c_0(r+s)}\frac{1}{2c_0}\left(1-e^{-2c_0(r\wedge s)}\right)\bar{h}_p(r,s)\,ds\,dr;
\end{aligned}
$$

and

$$
\begin{aligned}
&m_2(c)\\
={}& -(c-c_0)\left(\int_0^1\int_0^r e^{2c_0(r-s)}ds\,dr\right)\\
&+(c-c_0)\int_0^1\int_0^1\int_0^{r\wedge s} e^{c_0(r+s-2v)}\tilde{g}_{pc}(s)'A_p(c)^{-1}\tilde{g}_{pc}(r)\,dv\,ds\,dr\\
&+(c-c_0)\int_0^1\int_0^r e^{c_0(r-s)}\tilde{g}_{pc}(r)'A_p(c)^{-1}g_p(s)\,ds\,dr\\
&+(c-c_0)\int_0^1\int_0^r e^{c_0(r-s)}\tilde{g}_{pc}(s)'A_p(c)^{-1}g_p(r)\,ds\,dr\\
&-(c-c_0)^2\int_0^1\int_0^1\int_0^{r\wedge s} e^{c_0(r+s-2v)}\tilde{g}_{pc}(s)'A_p(c)^{-1}g_p(r)\,dv\,ds\,dr\\
&-(c-c_0)\int_0^1\int_0^r e^{c_0(r-s)}\tilde{g}_{pc}(r)'A_p(c)^{-1}B_p(c)A_p(c)^{-1}\tilde{g}_{pc}(s)\,ds\,dr\\
&-(c-c_0)\int_0^1\int_0^r e^{c_0(r-s)}\tilde{g}_{pc}(s)'A_p(c)^{-1}B_p(c)'A_p(c)^{-1}\tilde{g}_{pc}(r)\,ds\,dr\\
&+(c-c_0)^2\int_0^1\int_0^1\int_0^{r\wedge s} e^{c_0(r+s-2v)}\tilde{g}_{pc}(s)'A_p(c)^{-1}B_p(c)A_p(c)^{-1}\tilde{g}_{pc}(r)\,dv\,ds\,dr\\
&-\int_0^1\int_0^r e^{c_0(r-s)}\tilde{g}_{pc}(s)'A_p(c)^{-1}\tilde{g}_{pc}(r)\,ds\,dr\\
&+\int_0^1\int_0^r e^{c(r-s)}\tilde{g}_{pc}(s)'A_p(c)^{-1}\tilde{g}_{pc}(r)\,ds\,dr.
\end{aligned}
$$

The following lemma shows that the sample moment condition $M_{nT}(c)$ has a uniform limit in c.

Lemma 2 (Uniform Convergence) Under Assumptions 1-6,

$$M_{nT}(c) \to_p \mathcal{M}(c;c_0) \text{ uniformly in } c$$

as $(n,T\to\infty)$.

Assumption 7 As $(n,T\to\infty)$, $\hat{W}\to_p W$, where W is positive definite.

Notice by inspection that the uniform limit function $\mathcal{M}(c;c_0)$ is continuous on the compact parameter set C. Also, notice that $\mathcal{M}(c;c_0)=0$ at the true parameter $c=c_0$. In Appendix F, we prove numerically that $\mathcal{M}(c;c_0)=0$ only when $c=c_0$. Then, by a standard result (e.g., theorem 2.1 of Newey and McFadden (1994), the GMM estimator $\hat{c}$ is consistent for the true parameter c_0. Summarizing, we have the following theorem.

Theorem 1 (Consistency) Suppose that Assumptions 1-6 and Assumption 7 hold. Then, as $(n,T\to\infty)$,

$$\hat{c}\to_p c_0.$$

4.2 Limiting Distribution of the GMM Estimator when $c_0 < 0$

By inspection the objective function $Z_{nT}(c)$ is differentiable in c on the region $c \in (\bar{c}, 0)$; and it has right and left derivatives at $c = \bar{c}$ and 0; respectively. To derive the limit distribution of the GMM estimator, we employ an approach that approximates the objective function $Z_{nT}(c)$ uniformly in terms of a quadratic function in a shrinking neighborhood of the true parameter.

For this purpose, we define

$$dM_{nT}(c) = \frac{1}{n}\sum_{i=1} dm_{iT}(c);$$

where $dm_{iT}(c)$ denotes the derivative of $m_{iT}(c)$ with respect to c when $c \in (\bar{c}, 0)$ and the right and left derivatives when $c = \bar{c}$ and 0; respectively. By the mean value theorem, for $c \neq c_0$;

$$m_{iT}(c) = m_{iT}(c_0) + dm_{iT}(c_0)(c - c_0) + r_{iT}(c;c_0)(c - c_0);$$

where

$$\begin{aligned} r_{iT}(c;c_0) &= (r_{1iT}(c;c_0), r_{2iT}(c;c_0))'; \\ r_{kiT}(c;c_0) &= dm_{kiT}\left(c_k^+\right) - dm_{kiT}(c_0); \end{aligned}$$

and c_k^+ lies between c and c_0 for $k = 1; 2$:

Define

$$S_{nT} = dM_{nT}(c_0)'\hat{W}M_{nT}(c_0);$$

and

$$H_{nT} = dM_{nT}(c_0)'\hat{W}dM_{nT}(c_0):$$

Then, we can write

$$\begin{aligned} Z_{nT}(c) &= M_{nT}(c_0)'\hat{W}M_{nT}(c_0) + 2(c - c_0)S_{nT} + (c - c_0)^2 H_{nT} \\ &\quad + (c - c_0)R_{1nT}(c;c_0) + (c - c_0)^2 R_{2nT}(c;c_0); \end{aligned}$$

where

$$R_{1nT}(c;c_0) = 2M_{nT}(c_0)'\hat{W}\left(\frac{1}{n}\sum_{i=1} r_{iT}(c;c_0)\right);$$

and

$$\begin{aligned} R_{2nT}(c;c_0) &= 2dM_{nT}(c_0)'\hat{W}\left(\frac{1}{n}\sum_{i=1} r_{iT}(c;c_0)\right) \\ &\quad + \left(\frac{1}{n}\sum_{i=1} r_{iT}(c;c_0)\right)'\hat{W}\left(\frac{1}{n}\sum_{i=1} r_{iT}(c;c_0)\right): \end{aligned}$$

We now give some asymptotic results that are useful in establishing the limit distribution of $\hat{c}$:

Lemma 3 Suppose that Assumptions 1-6 hold. When the true parameter is c_0;

$$dM_{nT}(c) \to_p dM(c;c_0) \equiv \begin{pmatrix} dM_1(c;c_0) \\ dM_2(c;c_0) \end{pmatrix} \text{ uniformly in } c \text{ as } (n, T \to \infty)$$

for some continuous function $dM(c)$ with

$$dM_1(c_0; c_0) = -\omega_2(c_0) + \int_0^1\int_0^r e^{c_0(r-s)}(r-s)\bar{h}_p(r;s)\,ds\,dr;$$

and

$$
\begin{aligned}
& dM_2(c_0; c_0) \\
= {} & -\int_0^1\int_0^r e^{2c_0(r-s)}ds\,dr \\
& + \int_0^1\int_0^1\int_0^{r\wedge s} e^{c_0(r+s-2v)}\tilde{g}_{pc_0}(s)' A_p(c_0)^{-1}\tilde{g}_{pc_0}(r)\,dv\,ds\,dr \\
& + \int_0^1\int_0^r e^{c_0(r-s)}\tilde{g}_{pc_0}(r)' A_p(c_0)^{-1} g_p(s)\,ds\,dr \\
& + \int_0^1\int_0^r e^{c_0(r-s)}\tilde{g}_{pc_0}(s)' A_p(c_0)^{-1} g_p(r)\,ds\,dr \\
& - \int_0^1\int_0^r e^{c_0(r-s)}\tilde{g}_{pc_0}(r)' A_p(c_0)^{-1} B_p(c_0) A_p(c_0)^{-1}\tilde{g}_{pc_0}(s)\,ds\,dr \\
& - \int_0^1\int_0^r e^{c_0(r-s)}\tilde{g}_{pc_0}(s)' A_p(c_0)^{-1} B_p(c_0)' A_p(c_0)^{-1}\tilde{g}_{pc_0}(r)\,ds\,dr \\
& + \int_0^1\int_0^r (r-s)e^{c_0(r-s)}\tilde{g}_{pc_0}(r)' A_p(c_0)^{-1}\tilde{g}_{pc_0}(s)\,ds\,dr.
\end{aligned}
$$

Now we set $B_{nT} = \sqrt{n}$.

Lemma 4 Suppose that Assumptions 1-6 hold. Then, as $(n;T \to \infty)$ with $\frac{n}{T} \to 0$;

$$B_{nT}M_{nT}(c_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} m_{iT}(c_0) \Rightarrow N\left(0;\ \sigma^2 J'\Omega(c_0)J\right);$$

where $J = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 1 \end{pmatrix}'$ and Ω is defined in (45).

Remarks

(a) The proof is similar to that of Lemma 2 and is omitted.

(b) Figures (3) and (4) plot the graphs of $dM_1(c_0; c_0)$ in the cases of $\bar{g}_{1t} = (1; t)'$ and $\bar{g}_{2t} = \left(1; t; t^2\right)'$; respectively. What we verify from the graphs is that $dM_1(c_0; c_0) < 0$ for $c_0 < 0$. Therefore, $H_{nT} > 0$ for $c_0 < 0$.

Figure 3. Graph of $dM_1(c_0; c_0)$ when $\bar{g}_{1t} = (1; t)'$.

Figure 4. Graph of $dM_1(c_0; c_0)$ when $g_{2t} = \left(1; t; t^2\right)'$:

(c) According to Moon and Phillips (1999b), when $c_0 = 0$; it always holds that $dM_1(c_0; c_0) = 0$ for all polynomial trends $g_{pt} = (1; \ldots; t^p)'$: Also, for $c_0 = 0$; direct calculations show that $dM_2(c_0; c_0) = 0$ for $g_{1t} = t$ and $dM_2(c_0; c_0) = 0$ for $g_{2t} = \left(t; t^2\right)'$: Therefore, $H_{nT} \to_p 0$ when $c_0 = 0$, $g_{1t} = t$; and $g_{2t} = \left(t; t^2\right)'$:

Notice from Lemma 3 and the following remarks and by Assumption 7, that H_{nT} has a positive limit as $(n; T \to \infty)$ when $c_0 < 0$: Thus, $H_{nT}^{-1} = O_p(1)$. Then, we can write

$$
\begin{aligned}
& B_{nT}^2 Z_{nT}(c) \\
= \; & M_{nT}(c_0)' \hat{W} M_{nT}(c_0) - \frac{(B_{nT} S_{nT})^2}{H_{nT}} \\
& + H_{nT} \left(B_{nT}(c - c_0) - \frac{B_{nT} S_{nT}}{H_{nT}} \right)^2 \\
& + B_{nT}(c - c_0) B_{nT} R_{1nT}(c; c_0) + (B_{nT}(c - c_0))^2 R_{2nT}(c; c_0) : \qquad (11)
\end{aligned}
$$

Lemma 5 Under Assumptions 1-6 and Assumption 7, for every sequence $\gamma_{nT} \to 0$; we have

(a)

$$\sup_{c \in C : |c - c_0| \le \gamma_{nT}} |B_{nT} R_{1nT}(c; c_0)| = o_p(1)$$

and

(b)

$$\sup_{c \in C : |c - c_0| \le \gamma_{nT}} |R_{2nT}(c; c_0)| = o_p(1) :$$

Theorem 2 Suppose that Assumptions 1-6 and Assumption 7 hold. Then,

$$B_{nT}(\hat{c} - c_0) = O_p(1) :$$

Lemma 5 establishes that two remainder terms $B_{nT} R_{1nT}(c; c_0)$ and $R_{2nT}(c; c_0)$ converge in probability to zero uniformly in the shrinking neighborhood of the true parameter. Also, Theorem 2 shows that the GMM estimator is B_{nT} $(= \sqrt{n}) -$ consistent. This implies that in the shrinking neighborhood of the true parameter, the scaled objective function $B_{nT}^2 Z_{nT}(c)$ is uniformly approximated by the following quadratic function

$$
\begin{aligned}
& B_{nT}^2 Z_{q;nT}(c) \\
= \; & M_{nT}(c_0)' \hat{W} M_{nT}(c_0) - \frac{(B_{nT} S_{nT})^2}{H_{nT}} + H_{nT} \left(B_{nT}(c - c_0) - \frac{B_{nT} S_{nT}}{H_{nT}} \right)^2 :
\end{aligned}
$$

The heuristic ideas of the limit theory are as follows. Let $B_{nT}(\hat{c}_q - c_0) = \arg\max_{c \in C} B_{nT}^2 Z_{q,nT}(c)$. Then, we may expect that a maximizer of $B_{nT}^2 Z_{nT}(c)$ will be close to the maximizer of $B_{nT}^2 Z_{q,nT}(c)$, suggesting that the GMM estimator $B_{nT}(\hat{c} - c_0)$ will be close to

$$
\begin{aligned}
B_{nT}(\hat{c}_q - c_0) &= \frac{B_{nT}S_{nT}}{H_{nT}} \text{ if } \left\{ B_{nT}(\bar{c} - c_0) \le \frac{B_{nT}S_{nT}}{H_{nT}} \le -B_{nT}c_0 \right\} \\
&= B_{nT}(\bar{c} - c_0) \text{ if } \left\{ B_{nT}(\bar{c} - c_0) > \frac{B_{nT}S_{nT}}{H_{nT}} \right\} \\
&= -B_{nT}c_0 \text{ if } \left\{ \frac{B_{nT}S_{nT}}{H_{nT}} > -B_{nT}c_0 \right\}.
\end{aligned}
$$

Notice that $\frac{B_{nT}S_{nT}}{H_{nT}} = O_p(1)$ and recall that it is assumed that the true parameter $\bar{c} < c_0 < 0$. In this case, the probabilities of the events $\left\{ B_{nT}(\bar{c} - c_0) > \frac{B_{nT}S_{nT}}{H_{nT}} \right\}$ and $\left\{ \frac{B_{nT}S_{nT}}{H_{nT}} > -B_{nT}c_0 \right\}$ will be very small and the scaled and centred estimator $B_{nT}(\hat{c}_q - c_0)$ will therefore be close with high probability to the random variable

$$\hat{\lambda}_{nT} = \frac{B_{nT}S_{nT}}{H_{nT}}.$$

In view of Lemmas 3 and 4 and Assumption 7,

$$B_{nT}S_{nT} \Rightarrow S \overset{d}{=} N\left(0; \kappa^2 a^2 \left[dM(c_0;c_0)' W J' \Omega(c_0) J W dM(c_0;c_0) \right]\right)$$

and

$$H_{nT} \to_p H = \kappa^2 dM(c_0;c_0)' W dM(c_0;c_0) > 0$$

as $(n,T \to \infty)$ with $\frac{n}{T} \to 0$. Thus, when $c_0 \in C_0 / \{0\}$,

$$\hat{\lambda}_{nT} \Rightarrow \lambda \overset{d}{=} H^{-1}S \overset{\text{let}}{=} Z.$$

The proof of the following theorem verifies the heuristic arguments given above.

Theorem 3 Suppose that Assumptions 1-6 and Assumption 7 hold. Suppose that $c_0 \in C_0/\{0\}$ and $\hat{c}$ be the GMM estimator defined in (10). Then, as $(n,T \to \infty)$ with $\frac{n}{T} \to 0$,

$$\sqrt{n}(\hat{c} - c_0) \Rightarrow Z,$$

where
$$Z \overset{d}{=} N\left(0, \frac{a^2}{\kappa^2} \frac{dM(c_0;c_0)' W J' \Omega(c_0) J W dM(c_0;c_0)}{\left[dM(c_0;c_0)' W dM(c_0;c_0)\right]^2}\right).$$

Remarks

(a) When $c_0 \in C_0/\{0\}$ and $J'\Omega(c_0)J$ is invertible, the optimal weight matrix is found as

$$\hat{W}_{opt} = (J'\Omega(\hat{c})J)^{-1}.$$

The limiting distribution of $\sqrt{n}(\hat{c} - c_0)$ is then

$$\sqrt{n}(\hat{c} - c_0) \Rightarrow Z_{opt} \overset{d}{=} N\left(0, \frac{a^2}{\kappa^2 \left[dM(c_0;c_0)' W dM(c_0;c_0)\right]^2}\right). \tag{12}$$

(b) In Figures 5-6, we plot the graphs of the minimum eigenvalues of $J'\Omega(c_0)J$ as functions of c_0 when $g_{1t} = t$ and $g_{2t} = (t, t^2)'$. As we see through the graphs, $J'\Omega(c_0)J$ is positive definite except for the case of $c_0 = 0$ with $g_{1t} = t$.

Figure 6. Graph of the Minimum Eigenvalue of $J'\Omega(c_0)J$ When $g_{2t} = (t, t^2)'$.

4.3 Limiting Distribution of the GMM Estimator when $c_0 = 0$

An important special case of model 1 is when $c_0 = 0$. In this case, the time series components of y_{it} in (1) have a unit root (i.e., $\rho_0 = 1$) for all i. This section develops asymptotics for the GMM estimator when the true localizing parameter is zero, so throughout this section we set $c_0 = 0$. In this case, according to the Remark (c) below Lemma 4, the information from the moment conditions is zero because $H_{nT} \to_p 0$. We cannot then use a conventional quadratic approximation approach, as in the previous section, and need instead to employ a higher order approximation.

The model considered is

$$
\begin{aligned}
z_{it} &= \beta_{i1} t + y_{it} && (13)\\
y_{it} &= \rho_0 y_{it-1} + \varepsilon_{it}, && (14)
\end{aligned}
$$

where

$$\rho_0 = 1, \ \text{i.e.}, \ c_0 = 0.$$

In model (13)-(14) the panel data z_{it} is generated by a heterogeneous deterministic trend, $\beta_{i1} t$, and has a nonstationary time series component y_{it} with a unit root. The analysis here is restricted to the linear trend case because it is the most widely used deterministic specification in empirical application and it facilitates what a complex series of calculations. Assumptions 2, 3, 4(a), 5, 6, and 7 are taken to hold.

Lemma 6 Under the assumptions stated above, the following hold as $(n, T \to \infty)$ with $\frac{n}{T} \to 0$:

(a) $\sqrt{n} M_{1nT}(0) \Rightarrow N\left(0, \frac{\omega^2}{60}\right) \equiv \sqrt{\frac{\omega^2}{60}} Z$, where $Z \equiv N(0,1)$;

(b) $\sqrt{n}\, d M_{1nT}(0) = O_p(1)$;

(c) $\sqrt{n}\, d^2 M_{1nT}(0) = o(1)$;

(d) $d^3 M_{1nT}(c) \to_p d^3 M_1(c;0)$ uniformly in c with $d^3 M_1(0;0) = -\frac{1}{70}$, where $d^k M_{1nT}(c)$ is the k^{th} left derivative of $M_{1nT}(c)$, and $d^3 M_1(c;0)$ is the third left derivative of $M_1(c;0)$, the probability limit of $M_{1nT}(c)$.

The next lemma finds the limits of the second moment condition and its higher order derivatives at $c = 0$. As we will show in the appendix, the asymptotics of $M_{2nT}(0)$ depend on the limiting behavior of $\frac{1}{n}\sum_{i=1}^n \frac{1}{T}\sum_{t=1}^T \left(\hat u_{it}^2 - \varepsilon_{it}^2\right)$, which relies on how we estimate the model and define the residual $\hat u_{it}$. The residual $\hat u_{it}$ that will be used here is obtained from a modified least squares estimation of model (4). In particular, we define

$$\hat u_{it} = \underset{\sim}{z}_{it} - \hat\rho^{++}\left(\underset{\sim}{z}_{i,-1}\right)_t, \tag{15}$$

where

$$\hat\rho^{++} = \left(\sum_{i=1}^n \underset{\sim}{z}_{i,-1}'\underset{\sim}{z}_{i,-1}\right)^{-1}\left(\sum_{i=1}^n\left(\underset{\sim}{z}_i'\underset{\sim}{z}_{i,-1} - T\hat\Omega_i - T\hat\lambda_i\omega_{1T}(0)\right)\right). \tag{16}$$

Then, we have the following lemma.

Lemma 7 Suppose that the assumptions in Lemma 6 hold. Assume that the residual $\hat u_{it}$ in (15) is used in calculating $\hat\lambda_i$ and $\hat\Omega_i$ in Assumption 6. Then, when $(n, T \to \infty)$ with $\frac{n}{T} \to 0$,

(a) $\sqrt{n} M_{2nT}(0) = o_p(1)$;

(b) $\sqrt{n}\, d M_{2nT}(0) = O_p(1)$;

(c) $\sqrt{n}\, d^2 M_{2nT}(0) = o_p(1)$;

(d) $d^3 M_{2nT}(c) \to_p d^3 M_2(c;0)$ uniformly in c with $d^3 M_2(0;0) = -\frac{1}{15}$, where $d^k M_{2nT}(0)$ is the k^{th} left derivative of $M_{2nT}(c)$ at $c = 0$, and $d^3 M_2(0;0)$ is the third left derivative of $d^3 M_2(c;0)$ at $c = 0$.

Remarks. Since the higher order derivatives of $M_{2nT}(0)$ are complicated and involve very lengthy expressions, we omit the details of their derivation in the appendix. Instead, we give a sketch of the proof in the appendix and here provide some simulation evidence relating to the various parts of Lemmas 6 and 7. Using simulated data for z_{it} in (13) with $\varepsilon_{it} \sim \text{iid } N(0,1)$ and $y_{i0} = 0$, we estimate the means and the variances of $\sqrt{n}\, d^k M_{jnT}(0)$, $k = 0, \ldots, 2$, $j = 1, 2$ and the means of $d^3 M_{jnT}(0)$, $j = 1, 2$. Table 1 reports the results. The numbers in the table are consistent with the theoretical results in the lemmas. Noticeably, the variance estimates of $\sqrt{n} M_{1nT}(0)$, $\sqrt{n}\, d M_{1nT}(0)$, and $\sqrt{n}\, d M_{2nT}(0)$ are all small. This is because their theoretical limit variances are small but not zero. In fact, a long calculation shows that the theoretical limit variances of $\sqrt{n} M_{1nT}(0)$, $\sqrt{n}\, d M_{1nT}(0)$, and $\sqrt{n}\, d M_{2nT}(0)$ are $\frac{1}{60}$, $\frac{11}{6300}$, and $\frac{1}{45}$, respectively when $\varepsilon_{it} \sim \text{iid } N(0,1)$.

Table 1[3]

	$\sqrt{n}M_{1nT}(0)$	$\sqrt{n}dM_{1nT}(0)$	$\sqrt{n}d^2M_{1nT}(0)$	$d^3M_{1nT}(c)$
Mean	-0.0019	-0.0003	7.96×10^{-7}	-0.0169
Variance	0.018	-0.0017	0	0
	$\sqrt{n}M_{2nT}(0)$	$\sqrt{n}dM_{2nT}(0)$	$\sqrt{n}d^2M_{2nT}(0)$	$d^3M_{2nT}(c)$
Mean	9.4×10^{-5}	-0.0001	-2.88×10^{-6}	-0.06
Variance	0.0012	0.022	4.85×10^{-6}	4.039

Using the left derivatives of the moment condition $m_{iT}(c)$ at $c = 0$; we approximate $m_{iT}(c)$ around the true parameter $c_0 = 0$ with a third order polynomial as follows,

$$m_{iT}(c) = m_{iT}(0) + c\,(dm_{iT}(0)) + \frac{1}{2}c^2\left[d^2 m_{iT}(0)\right] + \frac{1}{6}c^3\left[d^3 m_{iT}(0)\right] + c^3 \bar{r}_{iT}(c;0);$$

where

$$\begin{aligned} \bar{r}_{iT}(c;0) &= (\bar{r}_{1iT}(c;0), \bar{r}_{2iT}(c;0))'; \\ \bar{r}_{kiT}(c;0) &= d^3 m_{kiT}\left[c_k^+\right] - d^3 m_{kiT}(0); \; k = 1 \text{ and } 2. \end{aligned}$$

Then,

$$\begin{aligned} Z_{nT}(c) &= M_{nT}(c)'\hat{W} M_{nT}(c) \\ &= \sum_{k=0} c^k A_{k;nT} + N_{nT}(c;0); \end{aligned}$$

where

$$\begin{aligned} A_{0;nT} &= M_{nT}(0)'\hat{W} M_{nT}(0); \\ A_{1;nT} &= 2M_{nT}(0)'\hat{W} dM_{nT}(0); \\ A_{2;nT} &= M_{nT}(0)'\hat{W} d^2 M_{nT}(0) + dM_{nT}(0)'\hat{W} dM_{nT}(0); \\ A_{3;nT} &= \frac{1}{3}M_{nT}(0)'\hat{W} d^3 M_{nT}(0) + dM_{nT}(0)'\hat{W} d^2 M_{nT}(0); \\ A_{4;nT} &= \frac{1}{3}M_{nT}(0)'\hat{W} d^3 M_{nT}(0) + \frac{1}{4}d^2 M_{nT}(0)'\hat{W} d^2 M_{nT}(0); \\ A_{5;nT} &= \frac{1}{6}d^2 M_{nT}(0)'\hat{W} d^3 M_{nT}(0); \\ A_{6;nT} &= \frac{1}{36}d^3 M_{nT}(0)'\hat{W} d^3 M_{nT}(0); \end{aligned}$$

and

$$N_{nT}(c;0) = \sum_{k=3} c^k N_{k;nT}(c;0);$$

[3] Notice that the second and the third derivatives of $M_{1nT}(c)$ are deterministic.

$$N_{k;nT}(c;0) = 2d^{(k-3)} M_{nT}(0)'\hat{W}\left(\frac{1}{n}\sum_{i=1} r_{iT}(c;0)\right) \quad \text{for } k = 3;4;5;^{4}$$

$$N_{6;nT}(c;0) = 2d^{3} M_{nT}(0)'\hat{W}\left(\frac{1}{n}\sum_{i=1} r_{iT}(c;0)\right) + \left(\frac{1}{n}\sum_{i=1} r_{iT}(c;0)\right)'\hat{W}\left(\frac{1}{n}\sum_{i=1} r_{iT}(c;0)\right):$$

In view of Lemmas 6 and 7, it is easy to find that as $(n;T \to \infty)$ with $\frac{n}{T} \to 0;$

$$n^{5/6}A_{1;nT} = o_p(1); \tag{17}$$

$$n^{2/3}A_{2;nT} = o_p(1); \tag{18}$$

$$n^{1/3}A_{4;nT} = o_p(1); \tag{19}$$

$$n^{1/6}A_{5;nT} = o_p(1); \tag{20}$$

and

$$A_{6;nT} \to_p \frac{\omega^2}{36}\left(\frac{W_{11}}{4900} + \frac{2W_{12}}{1050} + \frac{W_{22}}{225}\right) > 0; \tag{21}$$

$$n^{1/2}A_{3;nT} \Rightarrow A_3 Z; \tag{22}$$

$$nA_{0;nT} \Rightarrow A_0 Z^2; \tag{23}$$

where $Z \sim N(0;1)$ and $A_3 = -\frac{\omega}{3}\left[\frac{W_{11}}{70} + \frac{W_{12}}{15}\right]\sqrt{\frac{a^2}{60}}$ and $A_0 = W_{11}\frac{a^2}{60}:$

Also, using Lemmas 6 and 7 and following similar lines of proof to Lemma 5, we can show that

$$\sup_{c \in C: |c| \le \gamma_{nT}} \left| n^{(6-k)/6} N_{k;nT}(c;0) \right| = o_p(1); \tag{24}$$

for any sequence γ_{nT} tending to zero as $(n;T \to \infty)$: Then, we have the following limit theory for $\hat{c}$ at the origin.

Theorem 4 Under the assumptions in Lemmas 6 and 7, as $(n;T \to \infty)$ with $\frac{n}{T} \to 0;$

$$n^{1/6}(\hat{c} - c_0) = O_p(1);$$

where $c_0 = 0:$

So, when the true localizing parameter is $c_0 = 0;$ the GMM estimator $\hat{c}$ is $n^{1/6}-$consistent; which is slower than the regular case of $\sqrt{n}$ that applies for $c_0 < 0$ as shown in Section 4.

Next, we find the limiting distribution of the GMM estimator $\hat{c}$: The argument here is similar to that of the previous section. So, the proof is omitted and we give only the final result in Theorem 5 below.

In view of (17) $-$ (23) and (24); the standardized objective function $nZ_{nT}(c)$ is approximated by

$$Z_{q;nT}(c) = nA_{0;nT} + \left(n^{1/6}c\right)^3 \sqrt{n}A_{3;nT} + \left(n^{1/6}c\right)^6 A_{6;nT}:$$

Notice that the probability limit of $A_{6;nT}$ is positive, as shown in (21): Then, it is easy to see that the approximate objective function $Z_{q;nT}(c)$ is minimized at

$$\begin{aligned} n^{1/6}\hat{c}_q &= -\left(\frac{\sqrt{n}A_{3;nT}}{2A_{6;nT}}\right)^{1/3} \quad \text{if} \quad \left\{ n^{1/6}\hat{c} \le -\frac{\sqrt{n}A_{3;nT}}{2A_{6;nT}} \le 0 \right\} \\ &= 0 \text{ if } \left\{ -\frac{\sqrt{n}A_{3;nT}}{2A_{6;nT}} > 0 \right\} \\ &= -\left(n^{1/6}(-\hat{c})\right)^{1/3} \quad \text{if} \quad \left\{ n^{1/6}\hat{c} > -\frac{\sqrt{n}A_{3;nT}}{2A_{6;nT}} \right\}: \end{aligned}$$

Using arguments similar to those in the proof of Theorem 3, we can prove that the starndardized GMM estimator $n^{1/6}\hat{c}$ is approximated by $n^{1/6}\hat{c}_q$, the minimizer of $Z_{q,nT}(c)$; that is,

$$n^{1/6}\hat{c} = n^{1/6}\hat{c}_q + o_p(1),$$

and the estimator $n^{1/6}\hat{c}_q$ is approximated by

$$\hat{\lambda}_{nT} = -\left(\frac{\sqrt{n}A_{3,nT}}{2A_{6,nT}}\right)^{1/3} 1\left\{-\frac{\sqrt{n}A_{3,nT}}{2A_{6,nT}} \le 0\right\},$$

where $1\{A\}$ is the indicator of A. In view of (22) and (21), as $(n,T \to \infty)$ with $\frac{n}{T} \to 0$, it follows by the continuous mapping theorem that

$$\hat{\lambda}_{nT} \Rightarrow -(-Z_0)^{1/3} 1\{Z_0 \le 0\},$$

where

$$Z_0 = V_0 Z, \tag{25}$$

$$V_0 = \left| \frac{-\left(\frac{W_{11}}{70} + \frac{W_{12}}{15}\right)\sqrt{\frac{1}{15}}}{-\frac{1}{3}\left(\frac{W_{11}}{4900} + \frac{2W_{12}}{1050} + \frac{W_{22}}{225}\right)} \right| \tag{26}$$

and we have the following theorem.

Theorem 5 Under the assumptions in Lemmas 6 and 7, as $(n,T \to \infty)$ with $\frac{n}{T} \to 0$,

$$n^{1/6}\hat{c} \Rightarrow -(-Z_0)^{1/3} 1\{Z_0 \le 0\},$$

where Z_0 is defined in (25).

Remarks

(a) Theorem 4 shows that when the true parameter $c_0 = 0$, i.e., in the case of a panel unit root, the GMM estimator is $n^{1/6}$-consistent and that its limit distribution is nonstandard, involving the cube root of a truncated normal. The truncation in the limiting distribution arises because the true parameter is on the boundary of the parameter set.

(b) The reason for the slower convergence rate in the panel unit root case is that first order information in the moment condition (from the first derivative of the moment condition) is aymptotically zero at the true parameter. In order to obtain nonneglible information from the moment condition, we need to pass to third order derivatives of the moment condition. Taking the higher order approximation slows down the convergence rate because the rate at which information in the moment condition is passed to the estimator is slowed down at the origin because of the zero lower derivatives.

(c) In view of Lemmas 6(a) and 7(a), we find that $\sqrt{n}M_{2nT}(0) = o_p(1)$, while $\sqrt{n}M_{1nT}(0)$ converges in distribution to a normal random variable with positive variance. Because of the convergence rate difference between $\sqrt{n}M_{2nT}(0)$ and $\sqrt{n}M_{1nT}(0)$, we have only W_{11} and W_{12} but not W_{22} in the limiting scale V_0 of (26). In this case, setting $W_{11} = W_{12} = 0$, i.e. not considering the first moment condition, causes the variance of the limit variate Z_0 to vanish, from which one might expect that the GMM estimator from the second moment condition alone would have a faster convergence rate than $n^{1/6}$. In fact, under the assumptions in Lemma 7, it is possible to show that $nM_{2nT}(0) = o_p(1)$ as $(n, T \to \infty)$ with $\frac{n}{T} \to \infty$ and the GMM estimator from the second moment condition only could be $n^{1/4}$-consistent, which is faster than the GMM estimator defined by the two moment condition. However, the reason for using the first moment condition is to identify the true parameter when $c_0 < 0$. As we discuss in Appendix F, the second moment condition cannot identify the true parameter unless it is zero.

(d) When $c_0 = 0$, in view of Lemma 7(b) and (c), one can explore higher derivatives as moment conditions. If these higher derivative moment conditions are satisfied only at $c_0 = 0$, then it will be possible to use those moment conditions to distinguish the presence of a unit root in the panel from local alternatives, an issue which is being studied by the authors.

5 Monte Carlo Simulations

The purpose of this section is to compare the quantile dispersion of the GMM estimators in a simple simulation design. The main focus is to compare the panel unit root model with incidental trends with near unit root with incidental trends and panel unit root without the incidental trends.

The panel data z_{it} is generated by the system

$$
\begin{aligned}
z_{it} &= \beta_{i0} t + y_{it}, \qquad \beta_{i0} = \text{iid Uniform}[0, 3] \\
y_{it} &= \left(1 + \frac{c_0}{T}\right) y_{it-1} + \varepsilon_{it}, \qquad c_0 \in \{-20, -10, -5, 0\},
\end{aligned}
\tag{27}
$$

where the ε_{it} are iid $N(0, 1)$ across i and over t, and the initial values of y_{i0} are zeros. The sample size is $(n, T) = (100, 200)$. The autoregressive coefficients in the error process for y_{it} are taken to be 0.9, 0.95, 0.975, and 1. To calculate the GMM estimators we use an identity weight matrix. This choice makes the estimation procedure for the $c_0 < 0$ case comparable with the $c_0 = 0$ case, whereas the optimal weight matrix when $c_0 = 0$ is to use only the second moment condition in which case we can not identify the true parameter when $c_0 < 0$. The simulation employs 1000 repetitions each using grid search optimization with the grid length of 0.02.

The simulation results are reported in Table 2. First, the median bias of the GMM estimator $\hat{c}$ becomes larger as the true c_0 becomes larger. When $c_0 = 0$, the GMM estimator of Model (27) has median bias of -0.26, which is much larger than other cases. Also, when $c_0 = 0$, the GMM estimator is much more dispersed than the other cases. Both results are to be expected from the asymptotic theory because of the slower convergence rate and one sided limit distributin in the $c_0 = 0$ case.

Table 2 compares the GMM estimator in the panel unit root model with incidental trends with the truncated pooled OLS estimator of the panel unit root model without the trends. For this we calculate

$$
\hat{c} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} z_{it} z_{it-1}}{\sum_{i=1}^{n} \sum_{t=1}^{T} z_{it-1}^2} 1\left\{ \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} z_{it} z_{it-1}}{\sum_{i=1}^{n} \sum_{t=1}^{T} z_{it-1}^2} \le 0 \right\},
$$

where z_{it} is generated by Model (27) with $c_0 = 0$ and $\beta_{i0} = 0$. Then, the limting distribution of $\hat{c}$ is

$$
\begin{aligned}
\sqrt{n}\,\hat{c} &\Rightarrow \sqrt{2} Z 1\{Z \le 0\}, \\
Z &\sim N(0, 1),
\end{aligned}
$$

as $(n, T \to \infty)$; and so $\tilde{c}$ is $\sqrt{n}$- consistent and has a normal limiting distribution. The quantiles of $\tilde{c}$ when $n = 100$ and $T = 200$ are reported in the last row of Table 2. Comparing these outcomes with the GMM estimator $\hat{c}$ of Model (27) where incidental trends are present, $\tilde{c}$ is much more concentrated on the true value and the median bias of $\tilde{c}$ is much smaller than that of $\hat{c}$. This comparison highlights the delimiting effects of incidental trends on the estimation of roots near unity even in cases where there are long stretches of time series and cross section data in the panel.

Table 2. Quantiles of the Centered GMM Estimators of Model (27)

c_0 (ρ_0)	5%	10%	20%	30%	40%	50%	60%	70%	80%	90%	95%
-20 (0.9)	-1.38	-1.14	-0.82	-0.60	-0.38	-0.22	0	0.18	0.36	0.72	0.90
-10 (0.95)	-1.1	-0.86	-0.62	-0.44	-0.30	-0.16	0	0.16	0.34	0.54	0.80
-5 (0.975)	-0.92	-0.74	-0.52	-0.38	-0.24	-0.12	0	0.14	0.30	0.53	0.70
0 (1)	-1.64	-1.34	-0.96	-0.66	-0.42	-0.26	-0.1	0	0	0	0
0 (1) No Trend	-0.266	-0.197	-0.123	-0.075	-0.037	-0.003	0	0	0	0	0

6 Conclusion

Part of the richness of panel data is that it can provide information about features of a model on which time series and cross section data are uninformative when they are used on their own. In the context of nonstationary panels with near unit roots, an interesting new example of this 'added information' feature of panel data is that consistent estimation of the common local to unity coefficient becomes possible. This means that panel data help to sharpen our capacity to learn from data about the precise form of nonstationarity where time series data alone are insufficient to do so. However, as the authors have shown in earlier work, the presence of individual deterministic trends in a panel model introduces a serious complication in this nice result on the consistent estimation of a root local to unity. The complication is that individual trends produce an incidental parameter problem as $n \to \infty$ that does not disappear as $T \to \infty$. The outcome is that common procedures like pooled least squares and maximum likelihood are inconsistent. Thus, the presence of deterministic trends continues to confabulate inference about stochastic trends even in the panel data case.

One option is to adjust procedures like maximum likelihood to deal with the bias. The present paper shows how to make these adjustments. The theory is cast in the context of moment formulae that lead naturally to GMM based estimation. The paper has two important findings.

The first is that bias correction in the moment formulae arising from GLS estimation of the trend coefficients corresponds to taking the projected score (under Gaussian assumptions) on the Bhattacharya basis. This correspondence relates the approach we take here to recent work on projected score methods by Waterman and Lindsay (1998) that deals with models that have infinite numbers of nuisance parameters like the original incidental parameters problem.

The second is that our limit theory validates GMM-based inference about the localizing coefficient in near unit root panels. A notable new result is that the GMM estimator has a convergence rate slower than $\sqrt{n}$ when the true localizing parameter is zero (i.e., when there is a panel unit root) and the deterministic trends in the panel are linear. The asymptotic theory in this case provides a new example of limit theory on the boundary of a parameter space. The results point to the continued difficulty of distinguishing unit roots from local alternatives when there are deterministic trends in the data even when time series data is coupled with an infinity of additional data from a cross section.

7 Appendix

7.1 Appendix A:

Before we start the proof of Lemma 1, we give some useful background results.

Lemma 8 Let K_m denote the $(m \times m)$ commutation matrix, D_m denote the $m^2 \times \frac{1}{2}m(m+1)$ duplication matrix, and set $D_m^+ = (D_m' D_m)^{-1} D_m'$. Also, assume that x and y are m-vectors and A is an $(m \times m)$ invertible matrix. Then the following hold.

(a) $xy' \otimes yx' = K_m(yy' \otimes xx')$.

(b) $(I_m + K_m)((x \otimes y) + (y \otimes x)) = 2(x \otimes y) + 2(y \otimes x)$.

(c) $D_p^+ D_p = I_{\frac{1}{2}p(1+p)}$.

(d) $D_p D_p^+ = \frac{1}{2}(I_p + K_p)$.

(e) $\left(D_p^+(A \otimes A)D_p\right)^{-1} = D_p^+\left(A^{-1} \otimes A^{-1}\right)D_p$.

Proof

Parts (c), (d), and (e) are standard results (e.g., Magnus and Neudecker, 1988, pp. 49-50). Part (a) holds because

$$\begin{aligned} xy' \otimes yx' &= (x \otimes y)(y' \otimes x') = \mathrm{vec}(yx')(\mathrm{vec}(xy'))' \\ &= (K_m \mathrm{vec}(xy'))(\mathrm{vec}(xy'))' = K_m(y \otimes x)(y \otimes x)' \\ &= K_m(yy' \otimes xx'). \end{aligned}$$

Part (b) holds because

$$\begin{aligned} & (I_m + K_m)((x \otimes y) + (y \otimes x)) \\ = \; & (x \otimes y) + (y \otimes x) + K_m \mathrm{vec}(yx') + K_m \mathrm{vec}(xy') \\ = \; & (x \otimes y) + (y \otimes x) + \mathrm{vec}(xy') + \mathrm{vec}(yx') \\ = \; & 2(x \otimes y) + 2(y \otimes x). \; \blacksquare \end{aligned}$$

Proof of Lemma 1

In this proof we omit the subscript p that denotes the order of the polynomial trends for notational simplicity. To complete the proof, it is enough to show that $-\lambda_T(c)$ in $m_{2,iT}(c)$ is equivalent to $\lambda_2' D_p^+ \sum_{t=1}^T (\Delta_c g_t \otimes \Delta_c g_t)$ in $U_{2i}\left(c; \hat{\omega}_i(c)\right)$. First, we define

$$\begin{aligned} \hat{A}_{1T} &= \frac{1}{T}\sum_{t=2}^{T}\frac{1}{T}\sum_{s=1}^{t-1} D_p^+\left[\Delta_c\tilde{g}_t \otimes \Delta_c\tilde{g}_s + \Delta_c\tilde{g}_s \otimes \Delta_c\tilde{g}_t\right] e^{\left(\frac{t-s-1}{T}\right)c}; \\ \hat{A}_{2T} &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{T}\sum_{s=1}^{T} D_p^+\left\{\left(\Delta_c\tilde{g}_t\Delta_c\tilde{g}_t' \otimes \Delta_c\tilde{g}_s\Delta_c\tilde{g}_s'\right) + \left(\Delta_c\tilde{g}_t\Delta_c\tilde{g}_s' \otimes \Delta_c\tilde{g}_s\Delta_c\tilde{g}_t'\right)\right\}\left(D_p^+\right)'; \\ \hat{A}_{3T} &= D_p^+\frac{1}{T}\sum_{t=1}^{T}\left(\Delta_c\tilde{g}_t \otimes \Delta_c\tilde{g}_t\right). \end{aligned}$$

Then, by definition, we write

$$\lambda_2' D_p^+ \sum_{t=1}^{T}\left(\Delta_c g_t \otimes \Delta_c g_t\right) = \hat{A}_{1T}'\hat{A}_{2T}^{-1}\hat{A}_{3T}.$$

Notice by Lemma 8(a), (d), and (c) that

$$
\begin{aligned}
&\bar{A}_{2T}\\
=\;& D_p^+(I_p+K_p)\frac{1}{T}\sum_{t=1}^{T}\frac{1}{T}\sum_{s=1}^{T}\left(\Delta\tilde{g}_t\Delta\tilde{g}_t' - \Delta\hat{g}_s\Delta\tilde{g}_s'\right)\left(D_p^+\right)'\\
=\;& 2D_p^+D_pD_p^+\left[\left(\frac{1}{T}\sum_{t=1}^{T}\Delta\tilde{g}_t\Delta\hat{g}_t'\right) - \left(\frac{1}{T}\sum_{s=1}^{T}\Delta\hat{g}_s\Delta\tilde{g}_s'\right)\right]\left(D_p^+\right)'\\
=\;& 2D_p^+\left[\left(\frac{1}{T}\sum_{t=1}^{T}\Delta\tilde{g}_t\Delta\tilde{g}_t'\right) - \left(\frac{1}{T}\sum_{s=1}^{T}\Delta\hat{g}_s\Delta\tilde{g}_s'\right)\right]\left(D_p^+\right)'\\
=\;& 2\left[D_p^+\left(\frac{1}{T}\sum_{t=1}^{T}\Delta\tilde{g}_t\Delta\tilde{g}_t'\right) - \left(\frac{1}{T}\sum_{s=1}^{T}\Delta\hat{g}_s\Delta\tilde{g}_s'\right)D_p\right]\left(D_p'D_p\right)^{-1}.
\end{aligned}
$$

By Lemma 8(e),

$$
\begin{aligned}
&\bar{A}_{2T}^{-1}\\
=\;& \frac{1}{2}\left(D_p'D_p\right)D_p^+\left[\left(\frac{1}{T}\sum_{t=1}^{T}\Delta\tilde{g}_t\Delta\tilde{g}_t'\right)^{-1} - \left(\frac{1}{T}\sum_{s=1}^{T}\Delta\hat{g}_s\Delta\hat{g}_s'\right)^{-1}\right]D_p\\
=\;& \frac{1}{2}D_p'\left[\left(\frac{1}{T}\sum_{t=1}^{T}\Delta\tilde{g}_t\Delta\hat{g}_t'\right)^{-1} - \left(\frac{1}{T}\sum_{s=1}^{T}\Delta\tilde{g}_s\Delta\hat{g}_s'\right)^{-1}\right]D_p.
\end{aligned}
$$

Again, from Lemma 8(d) and (b), we have

$$
\begin{aligned}
&\bar{A}_{1T}'\bar{A}_{2T}^{-1}\bar{A}_{3T}\\
=\;& \frac{1}{T}\sum_{t=2}^{T}\frac{1}{T}\sum_{s=1}^{t-1}\left[\Delta\tilde{g}_t - \Delta\hat{g}_s + \Delta\hat{g}_s - \Delta\tilde{g}_t\right]' e^{\left(\frac{t-s-1}{T}\right)c}\left(D_p^+\right)'\\
&\times\frac{1}{2}D_p'\left[\left(\frac{1}{T}\sum_{t=1}^{T}\Delta\tilde{g}_t\Delta\hat{g}_t\right)^{-1} - \left(\frac{1}{T}\sum_{s=1}^{T}\Delta\tilde{g}_s\Delta\hat{g}_s'\right)^{-1}\right]D_p\\
&\times D_p^+\frac{1}{T}\sum_{t=1}^{T}\left(\Delta\hat{g}_t - \Delta\tilde{g}_t\right)\\
=\;& \frac{1}{8}\left[\frac{1}{T}\sum_{t=2}^{T}\frac{1}{T}\sum_{s=1}^{t-1}\left[\Delta\hat{g}_t - \Delta\hat{g}_s + \Delta\hat{g}_s - \Delta\tilde{g}_t\right]' e^{\left(\frac{t-s-1}{T}\right)c}\left(I_p+K_p\right)\right.\\
&\times\left[\left(\frac{1}{T}\sum_{t=1}^{T}\Delta\tilde{g}_t\Delta\hat{g}_t\right)^{-1} - \left(\frac{1}{T}\sum_{s=1}^{T}\Delta\tilde{g}_s\Delta\hat{g}_s'\right)^{-1}\right]\\
&\left.\times\left(I_p+K_p\right)\left(\frac{1}{T}\sum_{t=1}^{T}\left(\Delta\hat{g}_t - \Delta\tilde{g}_t\right)\right)\right]\\
=\;& \frac{1}{2}\left[\frac{1}{T}\sum_{t=2}^{T}\frac{1}{T}\sum_{s=1}^{t-1}\left[\Delta\hat{g}_t - \Delta\hat{g}_s + \Delta\hat{g}_s - \Delta\tilde{g}_t\right]' e^{\left(\frac{t-s-1}{T}\right)c}\right.\\
&\times\left[\left(\frac{1}{T}\sum_{t=1}^{T}\Delta\tilde{g}_t\Delta\hat{g}_t\right)^{-1} - \left(\frac{1}{T}\sum_{s=1}^{T}\Delta\tilde{g}_s\Delta\hat{g}_s'\right)^{-1}\right]
\end{aligned}
$$

$$\times\left[\frac{1}{T}\sum_{t=1}^{T}\left(\Delta\tilde{g}_t \otimes \Delta\tilde{g}_t\right)'\right]. \tag{28}$$

Expanding (28) yields

$$\frac{1}{T}\sum_{t=2}^{T}\frac{1}{T}\sum_{s=1}^{t-1}\frac{1}{T}\sum_{p=1}^{T} e^{\left(\frac{t-s-1}{T}\right)c}\left[\Delta\tilde{g}_s' A_{pT}^{-1}\Delta\tilde{g}_p\right]\left[\Delta\tilde{g}_p' A_{pT}^{-1}\Delta\tilde{g}_t\right]$$

$$= \frac{1}{T}\sum_{t=2}^{T}\frac{1}{T}\sum_{s=1}^{t-1} e^{\left(\frac{t-s-1}{T}\right)c}\Delta\tilde{g}_s' A_{pT}^{-1}\Delta\tilde{g}_t$$

$$= \mathrm{tr}\left(A_{pT}^{-1}\frac{1}{T}\sum_{t=2}^{T}\frac{1}{T}\sum_{s=1}^{t-1} e^{\left(\frac{t-s-1}{T}\right)c}\Delta\tilde{g}_t\Delta\tilde{g}_s'\right)$$

$$= -\lambda_T(c). \ \blacksquare$$

7.2 Appendix B: Useful Results for Joint Asymptotic Theories

This section consists of two subsections. The first subsection introduces some useful results for joint asymptotic theories. Many of these are modified versions of results developed in Phillips and Moon (1999) so we report them only briefly here. The second subsection introduces some useful results which will be used repeatedly in the following sections of the proofs for the results in the main text.

7.2.1 Appendix B1

The following two theorems provide convenient conditions to find the joint probability limit of double indexed processes.

Theorem 6 (Joint Probability Limits) Suppose the $(m \times 1)$ random vectors Y_{iT} are independent across $i = 1, \ldots, n$ for all T and integrable. Assume that $Y_{iT} \Rightarrow Y_i$ as $T \to \infty$ for all i. Let $X_{nT} = \frac{1}{n}\sum_{i=1}^{n} Y_{iT}$ and $X_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

(a) Let the following hold:

- (i) $\limsup_{n,T} \frac{1}{n}\sum_{i=1}^{n} E\|Y_{iT}\| < \infty$;
- (ii) $\limsup_{n,T} \frac{1}{n}\sum_{i=1}^{n} \|EY_{iT} - EY_i\| = 0$;
- (iii) $\limsup_{n,T} \frac{1}{n}\sum_{i=1}^{n} E\|Y_{iT}\| 1\{\|Y_{iT}\| > n\varepsilon\} = 0 \ \forall \varepsilon > 0$; and
- (iv) $\limsup_{n} \frac{1}{n}\sum_{i=1}^{n} E\|Y_i\| 1\{\|Y_i\| > n\varepsilon\} = 0 \ \forall \varepsilon > 0$.

(b) If $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} EY_i \ (:= \mu_X)$ exists and $X_n \to_p \mu_X$ as $n \to \infty$, then $X_{nT} \to_p \mu_X$ as $(n, T \to \infty)$.

Theorem 7 Suppose that $Y_{iT} = C_i Q_{iT}$, where the $(m \times 1)$ random vectors Q_{iT} are iid across $i = 1, \ldots, n$ for all T, and the C_i are $(m \times m)$ nonrandom matrices for all i. Assume that

(i) $Q_{iT} \Rightarrow Q_i$ as $T \to \infty$ for all i as $(n, T \to \infty)$,

(ii) $\|Q_{iT}\|$ is uniformly integrable in T for all i[5].

[5] That is,

$$\sup_T E\{\|Q_{iT}\| 1\{\|Q_{iT}\| > M\}\} \to 0$$

as $M \to \infty$.

(iii) $\sup_i \|C_i\| < \infty$; $\inf_i \|C_i\| > 0$; and $C = \lim_n \frac{1}{n}\sum_{i=1}^n C_i$.
Then $\frac{1}{n}\sum_{i=1}^n Y_{iT} \to_p CE(Q_i)$ as $(n, T \to \infty)$.

Theorem 8 (Joint Limit CLT for Scaled Variates) Suppose that $Y_{iT} = C_i Q_{iT}$, where the $(m \times 1)$ random vectors Q_{iT} are iid$(0, \Sigma_T)$ across $i = 1, \ldots, n$ for all T and the C_i are $(m \times m)$ nonzero and nonrandom matrices. Assume the following conditions hold:

(i) Let $\sigma_T^2 = \lambda_{\min}(\Sigma_T)$ and $\liminf_T \sigma_T^2 > 0$;

(ii) $\frac{\max_{i \le n} \|C_i\|^2}{\lambda_{\min}\left(\sum_{i=1}^n C_i C_i'\right)} = O\left(\frac{1}{n}\right)$ as $n \to \infty$;

(iii) $\|Q_{iT}\|^2$ are uniformly integrable in T,

(iv) $\lim_{n,T} \frac{1}{n}\sum_{i=1}^n C_i \Sigma_{nT} C_i' = \Omega > 0$.

Then,

$$X_{nT} = \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_{iT} \Rightarrow N(0, \Omega) \text{ as } n, T \to \infty.$$

7.2.2 Appendix B2

Suppose that the panel process y_{it} is generated by

$$y_{it} = \exp\left(\frac{c_0}{T}\right) y_{it-1} + \varepsilon_{it};$$

where ε_{it} satisfies Assumptions (2)-(5). Again, for notational simplicity, we omit the indices n and T in the notation y_{it}.

(a) A particularly useful tool in treating the linear process ε_{it} is the BN decomposition which decomposes the linear filter into long-run and transitory elements. Phillips and Solo (1992) give details of how this method can be used to derive a large number of limit results. Under Assumption 2, the linear process $\varepsilon_{i,t}$ is decomposed as

$$\varepsilon_{it} = C_i u_{it} + \tilde{\varepsilon}_{it-1} - \tilde{\varepsilon}_{it}, \tag{29}$$

where $\tilde{\varepsilon}_{i,t} = \sum_{j=0}^{\infty} \tilde{C}_{ij} u_{it-j}$; and $\tilde{C}_{ij} = \sum_{k=j+1}^{\infty} C_{ik}$. Under the summability condition (c) in Assumption 2,

$$|C_i| \le \sum_{j=0}^{\infty} \bar{C}_j < \infty \tag{30}$$

and

$$E\tilde{\varepsilon}_{it}^2 \le \left(\sum_{j=0}^{\infty} j\bar{C}_j\right)^2 \le \left(\sum_{j=0}^{\infty} j^b \bar{C}_j\right)^2 < \infty; \tag{31}$$

where $b \ge 1$ and $\bar{C}_j = \sup_i |C_{ij}|$ (see Phillips and Solo, 1992).

(b) Next, recall that

$$\tilde{h}_{pT}(t, s) = D_{pT}\tilde{g}_{pt}'\left(\frac{1}{T}\sum_{t=1}^{T} D_{pT}\tilde{g}_{pt}\tilde{g}_{pt}' D_{pT}\right)^{-1}\tilde{g}_{ps} D_{pT}.$$

It is easy to see that when $t = [Tr]$ and $s = [Tv]$, as $T \to \infty$

$$\phi_T(t,s) \to g_p'(r)\left(\int g_p g_p'\right)^{-1} g_p(v) = \phi_p(r,v)$$

uniformly in $(r,p) \in [0,1] \times [0,1]$. The following limit also holds

$$\sup_{1 \le t,s \le T} \phi_{pT}(t,s) \to \sup_{0 \le r,v \le 1} \phi_p(r,v). \tag{32}$$

(c) Using the BN decomposition of ε_{it}, we can decompose y_{it} into two terms - a long-run component of y_{it} and a transitory component. By virtue of the definition of y_{it},

$$y_{it} = \sum_{s=1}^{t} \exp\left(c_0 \frac{(t-s)}{T}\right) \varepsilon_{is} + \exp\left(c_0 \frac{t}{T}\right) y_{i0}.$$

Using the BN decomposition (29) of ε_{it}, we can decompose y_{it} as

$$y_{it} = C_i x_{it} + R_{it}, \tag{33}$$

where

$$\begin{aligned} x_{it} &= \sum_{s=1}^{t} \exp\left(c_0 \frac{(t-s)}{T}\right) u_{is} \\ \text{and } R_{it} &= \exp\left(c_0 \frac{(t-1)}{T}\right) \tilde{\varepsilon}_{i0} - \tilde{\varepsilon}_{it} \\ &\quad + \sum_{s=1}^{t} \exp\left(c_0 \frac{(t-s-1)}{T}\right) \tilde{\varepsilon}_{is}\left(1 - \exp\left(\frac{c_0}{T}\right)\right) + \exp\left(\frac{t}{T} c_0\right) y_{i0}. \end{aligned}$$

For notational simplicity we also omit the indices n and T in x_{it} and R_{it}. Let $x_{i0} = 0$ for all i.

Next we introduce bounds for the moments of some random variables that will be frequently used in the following proofs. Throughout the paper we use $\bar{K}$ as a generic constant independent of the localizing parameter c_{n0}. Let $t = [Tr]$. As $(n, T \to \infty)$

$$E\left(\frac{x_{it}^2}{T}\right) = \frac{1}{T} \sum_{s=1}^{t} \exp\left(2c_0 \frac{t-s}{T}\right) \to \int_0^r \exp((r-s)2c_0)\, ds < \bar{K}; \tag{34}$$

$$\frac{1}{T} \sum_{t=1}^{T} \sqrt{E\left(\frac{x_{it}^2}{T}\right)} = \frac{1}{T} \sum_{t=1}^{T} \sqrt{\frac{1}{T} \sum_{s=1}^{t} \exp\left(2c_0 \frac{t-s}{T}\right)} \to \int_0^1 \left(\int_0^r e^{(r-s)2c_0} ds\right)^{\frac{1}{2}} dr < \bar{K}; \tag{35}$$

and

$$\begin{aligned} &\lim_{n,T} \sup_{1 \le i \le n} \sup_{1 \le t \le T} ER_{it}^2 \\ &\le \lim_{n,T} \sup_{1 \le i \le n} \sup_{1 \le t \le T} 4 \left\{ \begin{array}{c} \exp\left(2c_0 \frac{t-1}{T}\right) \sup_i E\tilde{\varepsilon}_{i0}^2 + \sup_i E\tilde{\varepsilon}_{it}^2 \\ + \left(1 - \exp\left(\frac{c_0}{T}\right)\right)^2 \sum_{v=1}^{t} \sum_{s=1}^{t} \exp\left(c_0 \frac{t-1-v}{T}\right) \exp\left(c_0 \frac{t-1-s}{T}\right) \sup_i E(\tilde{\varepsilon}_{is}\tilde{\varepsilon}_{iv}) \\ + \exp\left(2c_0 \frac{t}{T}\right) \sup_i Ey_{i0}^2 \end{array} \right\} \end{aligned}$$

$$
\begin{aligned}
&\le \lim_{n,T} 4 \sup_{1\le i\le n} \sup_{1\le t\le T} \left\{ \begin{array}{l} \left(\exp\left(2c_0\frac{t-1}{T}\right)+1\right)\left(\sum_{j=0}^{\infty} j|\tilde{C}_j|\right)^2 \\ +\left(1-\exp\left(\frac{c_0}{T}\right)\right)^2\left(\sup_n \sup_{1\le v,s\le t}\exp\left(c_0\frac{2t-2-v-s}{T}\right)\right) \\ \times \sum_{v=1}^{t}\sum_{s=1}^{t}\sup_i |E(\varepsilon_{is}\varepsilon_{iv})| \end{array} \right\} \\
&\quad + \lim_{n,T} 4 \sup_{1\le t\le T} \exp\left(2c_0\frac{t}{T}\right) \sup_i E y_{i0}^2 \\
&\le 4\left(\sum_{j=0}^{\infty} j|\tilde{C}_j|\right)^2 \left\{ \lim_{n,T} \begin{array}{l} \sup_{1\le t\le T}\left(\exp\left(2c_0\frac{t-1}{T}\right)+1\right) \\ +T^2\left(1-\exp\left(\frac{c_0}{T}\right)\right)^2 \sup_{1\le v,s,t\le T}\exp\left(c_0\frac{2t-2-v-s}{T}\right)\sup_{1\le t\le T}\frac{t^2}{T^2} \end{array} \right\} \\
&\quad +4\lim_{n,T}\sup_{1\le t\le T}\exp\left(2c_0\frac{t}{T}\right)\sup_i \sigma_{i0}^2 \\
&\le 4\left(\sum_{j=0}^{\infty} j|\tilde{C}_j|\right)^2 \left\{2+\bar{c}^2 e^{-3\bar{c}}\right\} + 4\sup_i \sigma_{i0}^2; \text{ because } C=[\bar{c},0] \\
&\le \tilde{K}; \qquad (36)
\end{aligned}
$$

where $\sigma_{i0}^2 = E\left(y_{i0}^2\right)$.

Lemma 9 Assume that, for $k = 1,\dots,K$, $h_k(c,\tilde{c})$ is a real-valued continuous function on the product of the parameter set $C \times C$ with $h_k(c,c) = 0$, and $l_k(x,y)$ is a real-valued continuous function on $[0,1] \times [0,1]$. Also, assume that $f(x,c)$ and $g(x,c)$ are continuously differentiable functions from $[0,1] \times C$ to R such that $f(x,c)g(y,c) - f(x,\tilde{c})g(y,\tilde{c}) = \sum_{k=1}^{K} h_k(c,\tilde{c})\, l_k(x,y)$. Suppose that $y_{it} = \exp\left(\frac{c_0}{T}\right) y_{it-1} + \varepsilon_{it}$, where ε_{it} follows Assumption 2. Assume that Assumption 3 holds for the initial condition y_{i0} and Assumption 5 holds for the cross sectional limit of the long-run variances. Then, as $(n,T \to \infty)$, the following hold.

(a) $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T} y_{it-1}^2 \to_p \Omega \int_0^1\int_0^r e^{2c_0(r-s)}\,ds\,dr$.

(b) $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it} f\left(\frac{t}{T},c\right)\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{it-1} g\left(\frac{t}{T},c\right) \to_p \Omega\int_0^1\int_0^r e^{c_0(r-s)} g(r,c) f(s,c)\,ds\,dr$ uniformly in c.

(c) $\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{it-1} f\left(\frac{t}{T},c\right)\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{it-1} g\left(\frac{t}{T},c\right)\right)$
$\to_p \Omega\int_0^1\int_0^1 f(r,c)g(s,c)\left(\int_0^{r\wedge s} e^{c_0(r+s-2v)}\,dv\right)ds\,dr$ uniformly in c.

(d) $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it} f\left(\frac{t}{T},c\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it} g\left(\frac{t}{T},c\right) \to_p \Omega\int_0^1\int_0^1 f(r,c)g(s,c)\,ds\,dr$ uniformly in c.

Proof

Part (a) From the decomposition (33), we write

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T} y_{it-1}^2 \\
=\;& \frac{1}{n}\sum_{i=1}^{n} C_i^2 \frac{1}{T^2}\sum_{t=2}^{T} x_{it-1}^2 + 2\frac{1}{n}\sum_{i=1}^{n} C_i \frac{1}{T^2}\sum_{t=2}^{T} x_{it-1}R_{it-1} + \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=2}^{T} R_{it-1}^2 + \frac{1}{n}\sum_{i=1}^{n}\frac{y_{i0}^2}{T^2} \\
=\;& I_a + 2II_a + III_a + IV_a, \text{ say.}
\end{aligned}
$$

Since $\sup_i E y_{i0}^2 < \infty$, $IV_a \to 0$ as $(n,T \to \infty)$. In what follows we show that $I_a \to_p \Omega\int_0^1\int_0^r e^{2c_0(r-s)}\,ds\,dr$ and $II_a, III_a \to_p 0$ as $(n,T \to \infty)$.

For I_a; recall that

$$I_a = \frac{1}{n}\sum_{i=1} C_i^2 \frac{1}{T^2}\sum_{t=2} x_{it-1}^2.$$

Define $Q_{inT} = \frac{1}{T^2}\sum_{t=2}^{T} x_{it-1}^2$. Note that $\{Q_{iT}\}_{i=1,\ldots,n}$ are iid across i. Since

$$T^{-\frac{1}{2}} x_{it} \Rightarrow J_{c_0;i}(r) = \int_0^r e^{c_0(r-s)} dW_i(s) \tag{37}$$

as $T \to \infty$ (see Phillips, 1987); where W_i is standard Brownian motion, we have by the continuous mapping theorem as $(n, T \to \infty)$,

$$Q_{iT} \Rightarrow Q_i = \int_0^1 J_{c_0;i}^2(r)dr. \tag{38}$$

Also, as $T \to \infty$ for fixed n,

$$Q_{iT} \Rightarrow Q_i = \int_0^1 J_{c_{n0};i}^2(r)dr. \tag{39}$$

Notice that $EQ_i = \int_0^1\int_0^r e^{2c_0(r-s)}dsdr$.

We will claim $I_a \to_p \phi \int_0^1\int_0^r e^{2c_0(r-s)}dsdr$ in joint limits as $(n, T \to \infty)$ by verifying conditions (i) - (iii) in Theorem 7. Condition (iv) holds because it is assumed in Assumption 2 that $\lim_n \frac{1}{n}\sum_{i=1}^n C_i^2 = \phi$ and $\inf_i |C_i| > 0$, and under Assumption 2, it holds $\sup_i |C_i| < \infty$. Condition (i) is obvious in view of (38) and (39). For condition (ii), observe that

$$\begin{aligned} EQ_{iT} &= \frac{1}{T}\sum_{t=2}\frac{1}{T}\sum_{s=1}\exp\left(\frac{t-s}{T}2c_0\right) \\ &\to \int_0^1\int_0^r e^{(r-s)2c_0}dsdr = EQ_i \text{ as } (n, T \to \infty). \end{aligned}$$

Since $Q_{iT}\ (\geq 0) \Rightarrow Q_i$ with $EQ_{iT} \to EQ_i$ as $(n, T \to \infty)$, $\{Q_{iT}\}_T$ are uniformly integrable in T by Theorem 5.4 in Billingsley (1968).

Next, we prove that

$$II_a = \frac{1}{n}\sum_{i=1} C_i \frac{1}{T^2}\sum_{t=2} x_{it-1}R_{it-1} \to_p 0;$$

and

$$III_a = \frac{1}{n}\sum_{i=1}\frac{1}{T^2}\sum_{t=2} R_{it-1}^2 \to_p 0 \text{ as } n, T \to \infty;$$

by showing that $E|II_a|, E|III_a| \to 0$ as $n, T \to \infty$.

First, we have

$$\begin{aligned} E|II_a| &= E\left|\frac{1}{n}\sum_{i=1} C_i\frac{1}{T^2}\sum_{t=2} x_{it-1}R_{it-1}\right| \\ &\leq \frac{1}{n}\sum_{i=1}|C_i| E\left|\frac{1}{T^2}\sum_{t=2} x_{it-1}R_{it-1}\right| \leq C\left\{\frac{1}{n}\sum_{i=1} E\left|\frac{1}{T^2}\sum_{t=2} x_{i,t-1}R_{i,t-1}\right|\right\}. \end{aligned}$$

Observe that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} E\left|\frac{1}{T^2}\sum_{t=2}^{T} x_{it-1}R_{it-1}\right| \\
\le\; & \frac{1}{\sqrt{T}}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T} E\left|\frac{x_{it-1}}{\sqrt{T}}R_{it-1}\right| \\
\le\; & \frac{1}{\sqrt{T}}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\sqrt{E\left|\frac{x_{it-1}}{\sqrt{T}}\right|^2 E|R_{it-1}|^2} = O\left(\frac{1}{\sqrt{T}}\right),
\end{aligned}
$$

where the equality holds by (35) and (36). Similarly, we can show that $III_a \to_p 0$ as $(n,T\to\infty)$ by proving that $E|III_a|\to 0$ as $(n,T\to\infty)$. Therefore we have all the required results to complete the proof of part (a). $\blacksquare$

Part (b) Using the BN-decomposition in (33), we write

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it} f\left(\frac{t}{T},c\right)\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{it-1} g\left(\frac{t}{T},c\right)\right) \\
=\;& I_b + II_b + III_b + IV_b,
\end{aligned}
$$

where

$$
\begin{aligned}
I_b &= \frac{1}{n}\sum_{i=1}^{n}\gamma_i\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{it} f\left(\frac{t}{T},c\right)\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} x_{it-1} g\left(\frac{t}{T},c\right)\right), \\
II_b &= \frac{1}{n}\sum_{i=1}^{n}C_i\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{\varepsilon}_{it-1}-\tilde{\varepsilon}_{it}) f\left(\frac{t}{T},c\right)\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} x_{it-1} g\left(\frac{t}{T},c\right)\right), \\
III_b &= \frac{1}{n}\sum_{i=1}^{n}C_i\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{it} f\left(\frac{t}{T},c\right)\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} R_{it-1} g\left(\frac{t}{T},c\right)\right), \\
IV_b &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{\varepsilon}_{it-1}-\tilde{\varepsilon}_{it}) f\left(\frac{t}{T},c\right)\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} R_{it-1} g\left(\frac{t}{T},c\right)\right).
\end{aligned}
$$

We will show that

$$
I_b \to_p \gamma\int_0^1\int_0^r e^{c_0(r-s)} g(r,c) f(s,c)\,ds\,dr \text{ uniformly in } c
$$

and

$$
II_b, III_b, IV_b \to_p 0 \text{ uniformly in } c
$$

as $(n,T\to\infty)$.

First, we establish Part (b) for fixed c (pointwise convergence). Now, as in Part (a), we apply Theorem 7. Let

$$
\begin{aligned}
Q_{iT}(c) &= \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{it} f\left(\frac{t}{T},c\right)\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} x_{it-1} g\left(\frac{t}{T},c\right)\right), \\
\text{and } Q_i(c) &= \left(\int_0^1 f(s,c)\,dW_i(s)\right)\left(\int_0^1 g(r,c)\,J_{c_0,i}(r)\,dr\right).
\end{aligned}
$$

Using (37) and the continuous mapping theorem, we can show that

$$Q_{iT}(c) \Rightarrow Q_i(c) \tag{40}$$

as $T \to \infty$ for fixed n and c, which verifies condition (i) in Theorem 7. Condition (ii) holds because it is assumed in Assumption 2 that $\lim_n \frac{1}{n}\sum_{i=1}^n \Omega_i \left(= C_i^2\right) = \Omega$ and $\inf_i |C_i| > 0$, and under Assumption 2, it holds $\sup_i |C_i| < \infty$. Condition (iii) holds for fixed c if

$$Q_{1iT}(c) = \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it} f\left(\frac{t}{T};c\right)\right)^2$$

and

$$Q_{2iT}(c) = \left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} x_{it-1} g\left(\frac{t}{T};c\right)\right)^2$$

are uniformly integrable in T for fixed c. Notice that $Q_{1iT}(c) \Rightarrow Q_{1i}(c) = \left(\int_0^1 f(r;c)dW_i(r)\right)^2 > 0$, and $EQ_{1iT}(c) = \frac{1}{T}\sum_{t=1}^{T} f\left(\frac{t}{T};c\right)^2 \to \int_0^1 f(r;c)^2 dr = EQ_{1i}(c)$ as $T \to \infty$ for all i. By Theorem 5.4 in Billingsley (1968), it follows that $Q_{1iT}(c)$ are uniformly integrable in T for fixed c. In a similar fashion, $Q_{2iT}(c)$ is also uniformly integrable in T for fixed c. Therfore, as $(n, T \to \infty)$,

$$I_b \to_p \Omega \int_0^1\int_0^r e^{c_0(r-s)} g(r;c) f(s;c)\, ds\, dr \text{ for fixed } c.$$

Next, define $X_{nT}(c) = \frac{1}{n}\sum_{i=1}^n Q_{iT}(c)$. To complete the proof, we need to show that $X_{nT}(c)$ is stochastically equicontinuous. That is, for given $\varepsilon > 0$ and $\eta > 0$, there exists $\delta > 0$ such that

$$\limsup_{(n,T\to\infty)} P\left\{\sup_{|c-\tilde{c}|<\delta;\, c,\tilde{c}\in C} |X_{nT}(c) - X_{nT}(\tilde{c})| > \varepsilon\right\} < \eta.$$

Then, since the parameter set C is compact, the pointwise convergence of $X_{nT}(c)$ and the stochastic equicontinuity of $X_{nT}(c)$ imply uniform convergence.

Now we show the stochastic equicontinuity of $X_{nT}(c)$. First, notice that

$$\begin{aligned}
&\sup_{|c-\tilde{c}|<\delta;\, c,\tilde{c}\in C} |X_{nT}(c) - X_{nT}(\tilde{c})| \\
&= \sup_{|c-\tilde{c}|<\delta;\, c,\tilde{c}\in C} \left|\frac{1}{n}\sum_{i=1}^n \frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T u_{it}x_{is-1}\left\{f\left(\frac{t}{T};c\right)g\left(\frac{s}{T};c\right) - f\left(\frac{t}{T};\tilde{c}\right)g\left(\frac{s}{T};\tilde{c}\right)\right\}\right| \\
&= \sup_{|c-\tilde{c}|<\delta;\, c,\tilde{c}\in C} \left|\frac{1}{n}\sum_{i=1}^n \frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T u_{it}x_{is-1}\left\{\sum_{k=1}^K h_k(c;\tilde{c})\, l_k\left(\frac{t}{T};\frac{s}{T}\right)\right\}\right| \\
&\le \sup_{1\le k\le K}\ \sup_{|c-\tilde{c}|<\delta;\, c,\tilde{c}\in C} |h_k(c;\tilde{c})| \frac{1}{n}\sum_{i=1}^n \left|\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T u_{it}x_{is-1}\left\{\sum_{k=1}^K l_k\left(\frac{t}{T};\frac{s}{T}\right)\right\}\right|.
\end{aligned}$$

Since $h_k(c;\tilde{c})$ is continuous on the compact set with $h_k(c;c) = 0$ for all $k = 1,\ldots,K$, we can make $\sup_{1\le k\le K}\sup_{|c-\tilde{c}|<\delta;c,\tilde{c}\in C} |h_k(c;\tilde{c})|$ arbitrarily small by choosing a small $\delta > 0$. Also, under the assumptions in the lemma, it is not difficult to show that $\frac{1}{n}\sum_{i=1}^n \left|\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T u_{it}x_{is-1}\left\{\sum_{k=1}^K l_k\left(\frac{t}{T};\frac{s}{T}\right)\right\}\right| = O_p(1)$. Therefore, $X_{nT}(c)$ is stochastically equicontinuous, and $I_b \to_p \Omega\int_0^1\int_0^r e^{c_0(r-s)} g(r;c)f(s;c)\,ds\,dr$ uniformly in c.

Next, for II_b notice that

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\tilde{\varepsilon}_{it-1}-\tilde{\varepsilon}_{it}\right) f\left(\frac{t}{T};c\right)\\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\tilde{\varepsilon}_{it}\left(f\left(\frac{t+1}{T};c\right)-f\left(\frac{t}{T};c\right)\right)+\frac{1}{\sqrt{T}}\tilde{\varepsilon}_{i0} f\left(\frac{1}{T};c\right)-\frac{1}{\sqrt{T}}\tilde{\varepsilon}_{iT} f(1;c)\\
&= \frac{1}{T\sqrt{T}}\sum_{t=1}^{T-1}\tilde{\varepsilon}_{it}\frac{f\left(\frac{t+1}{T};c\right)-f\left(\frac{t}{T};c\right)}{\frac{1}{T}}+\frac{1}{\sqrt{T}}\tilde{\varepsilon}_{i0} f\left(\frac{1}{T};c\right)-\frac{1}{\sqrt{T}}\tilde{\varepsilon}_{iT} f(1;c).
\end{aligned}
$$

For $II_b \to_p 0$ uniformly in c if we show that $E\sup_{c\in C}|II_b| \to 0$ as $(n,T\to\infty)$. Let $\sup_i C_i=\bar{C}$. Under Assumption 2, $\bar{C}$ is finite. Now

$$
\begin{aligned}
&E\sup_{c\in C}|II_b|\\
&\leq \bar{C}\sup_i E\sup_{c\in C}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\tilde{\varepsilon}_{it-1}-\tilde{\varepsilon}_{it}\right) f\left(\frac{t}{T};c\right)\right|\left|\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} x_{it-1}g\left(\frac{t}{T};c\right)\right|\\
&\leq \bar{C}\sup_i E\sup_{c\in C}\left|\frac{1}{T\sqrt{T}}\sum_{t=1}^{T-1}\tilde{\varepsilon}_{it}\frac{f\left(\frac{t+1}{T};c\right)-f\left(\frac{t}{T};c\right)}{\frac{1}{T}}\right|\left|\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} x_{it-1}g\left(\frac{t}{T};c\right)\right|\\
&\quad+\bar{C}\sup_i E\sup_{c\in C}\left|\frac{1}{\sqrt{T}}\tilde{\varepsilon}_{i0} f\left(\frac{1}{T};c\right)\right|\left|\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} x_{it-1}g\left(\frac{t}{T};c\right)\right|\\
&\quad+\bar{C}\sup_i E\sup_{c\in C}\left|\frac{1}{\sqrt{T}}\tilde{\varepsilon}_{iT} f(1;c)\right|\left|\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} x_{it-1}g\left(\frac{t}{T};c\right)\right|. \qquad (41)
\end{aligned}
$$

The first term on the RHS of (41) is less than or equal to

$$
\bar{C}\left(\sup_{\substack{1\leq t\leq T\\ c\in C}}\left|\frac{f\left(\frac{t+1}{T};c\right)-f\left(\frac{t}{T};c\right)}{\frac{1}{T}}\right|\right)\left(\sup_{\substack{1\leq t\leq T\\ c\in C}}\left|g\left(\frac{t}{T};c\right)\right|\right)
\times\sup_i E\left(\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T-1}|\tilde{\varepsilon}_{it}|\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}|x_{it-1}|\right)\right).
$$

Since $f(x;c)$ and $g(x;c)$ are continuously differentiable functions on the compact set $[0,1]\times C$, $\sup_{\substack{1\leq t\leq T\\ c\in C}}\left|\frac{f\left(\frac{t+1}{T};c\right)-f\left(\frac{t}{T};c\right)}{\frac{1}{T}}\right|$ and $\sup_{\substack{1\leq t\leq T\\ c\in C}}\left|g\left(\frac{t}{T};c\right)\right|$ are bounded by a constant, say $\bar{K}$, that is independent of c. Also,

$$
\begin{aligned}
&\sup_i E\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T-1}|\tilde{\varepsilon}_{it}|\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}|x_{it-1}|\right)\\
&\leq \frac{1}{\sqrt{T}}\sup_i\sqrt{E\left(\frac{1}{T}\sum_{t=1}^{T-1}|\tilde{\varepsilon}_{it}|\right)^2}\sqrt{E\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}|x_{it-1}|\right)^2}\quad\text{by Cauchy-Schwarz inequality}\\
&\leq \frac{1}{\sqrt{T}}\sup_i\sqrt{\frac{1}{T}\sum_{t=1}^{T-1}E\tilde{\varepsilon}^2_{it}}\sqrt{\frac{1}{T}\sum_{t=1}^{T}E\left(\frac{x_{it-1}}{\sqrt{T}}\right)^2}\\
&\leq \frac{\bar{K}}{\sqrt{T}}\quad\text{for some constant }\bar{K}\text{ independent of }c\text{ by (31) and (35).}
\end{aligned}
$$

Similarly, we can show that the other terms in the RHS of (41) are less than equal to $\frac{\bar K}{\sqrt{T}}$ for some constant $\bar K$ independent of c. Therefore

$$E\sup_{c\in C}|II_b| \le \frac{\bar K}{\sqrt{T}} \text{ for some constant } \bar K \text{ independent of } c, \tag{42}$$

and so $II_b \to_p 0$ uniformly in c.

In a similar fashion, it is possible to show that

$$E\sup_{c\in C}|III_b|,\; E\sup_{c\in C}|IV_b| \le \frac{\bar K}{\sqrt{T}} \text{ for some constant } \bar K \text{ independent of } c, \tag{43}$$

which leads to $III_b, IV_b \to_p 0$ uniformly in c. We omit the details of the argument here. ■

Part (c) and Part (d) The proofs of Parts (c) and (d) are similar to that of Part (b) and they are omitted. ■

The following lemma is important in establishing asymptotic normality of the GMM estimator $\hat c$. To simplify notation, let

$$\begin{aligned}
l_{1pT}(t,s;c) &= \Delta_c\tilde g_{pt}{}' A_{pT}(c)^{-1}\Delta_c\tilde g_{ps} \\
l_{2pT}(t,s;c) &= \Delta_c\tilde g_{pt}{}' A_{pT}(c)^{-1} g_{ps-1} D_T^{-1} \\
l_{3pT}(t,s;c) &= \Delta_c\tilde g_{pt}{}' A_{pT}(c)^{-1} B_{pT}(c) A_{pT}(c)^{-1}\Delta_c\tilde g_{ps},
\end{aligned}$$

and

$$\begin{aligned}
l_{1p}(r,s;c) &= \overset{\Delta}{g}_{pc}(r)' A_p(c)^{-1}\overset{\Delta}{g}_{pc}(s) \\
l_{2p}(r,s;c) &= \overset{\Delta}{g}_{pc}(r)' A_p(c)^{-1} g_p(s) \\
l_{3p}(r,s;c) &= \overset{\Delta}{g}_{pc}(r)' A_p(c)^{-1} B_p(c) A_p(c)^{-1}\overset{\Delta}{g}_{pc}(s) \\
l_{4p} &= \int_0^1 g_p(r) g_p(r)' dr.
\end{aligned}$$

Lemma 10 Suppose that $x_{it} = \exp\left(-\frac{c_0}{T}\right)x_{it-1} + u_{it}$, where u_{it} are iid$(0,1)$ with finite fourth moments and $x_{i0} = 0$ for all i. Then, as $(n,T \to \infty)$, the following hold.

Let

$$\begin{aligned}
Q_{1iT} &= \frac{1}{T}\sum_{t=1}^{T} x_{it-1}u_{it} \\
Q_{2iT} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{T\sqrt{T}}\sum_{s=1}^{T} u_{it}x_{is-1}\tilde h_{pT}(t,s) + \omega_{1T}(c_0) \\
Q_{3iT} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{T\sqrt{T}}\sum_{s=1}^{T} u_{it}x_{is-1}l_{1pT}(t,s;c_0) + \lambda_T(c_0) \\
Q_{4iT} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{\sqrt{T}}\sum_{s=1}^{T} u_{it}u_{is}l_{2pT}(t,s;c_0) - \operatorname{tr}\left(A_{pT}(c_0)^{-1}B_p(c_0)\right) \\
Q_{5iT} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{\sqrt{T}}\sum_{s=1}^{T} u_{it}u_{is}l_{3pT}(t,s;c_0) - \operatorname{tr}\left(A_{pT}(c_0)^{-1}B_p(c_0)\right) \\
\text{and } Q_{iT} &= (Q_{1iT}, Q_{2iT}, Q_{3iT}, Q_{4iT}, Q_{5iT})'.
\end{aligned} \tag{44}$$

Then, as $(n, T \to \infty)$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_i Q_{iT} \Rightarrow N\left(0, \sigma^2 \Phi(c_0)\right),$$

where

$$\Phi(c_0) = \begin{bmatrix} \Phi_{11}(c_0) & \Phi_{12}(c_0) & \Phi_{13}(c_0) & \Phi_{14}(c_0) & \Phi_{15}(c_0) \\ \Phi_{12}(c_0) & \Phi_{22}(c_0) & \Phi_{23}(c_0) & \Phi_{24}(c_0) & \Phi_{25}(c_0) \\ \Phi_{13}(c_0) & \Phi_{23}(c_0) & \Phi_{33}(c_0) & \Phi_{34}(c_0) & \Phi_{35}(c_0) \\ \Phi_{14}(c_0) & \Phi_{24}(c_0) & \Phi_{34}(c_0) & \Phi_{44}(c_0) & \Phi_{45}(c_0) \\ \Phi_{15}(c_0) & \Phi_{25}(c_0) & \Phi_{35}(c_0) & \Phi_{45}(c_0) & \Phi_{55}(c_0) \end{bmatrix} \tag{45}$$

and

$$\Phi_{11}(c_0) = \int_0^\infty \int_0^r e^{2c_0(r-s)} ds dr;$$

$$\Phi_{12}(c_0) = \int_0^\infty \int_0^\infty \int_0^{r\wedge s} e^{c_0(r+s-2v)} \rho_p(r,s) dv ds dr + \int_0^\infty \int_0^r \int_0^s e^{c_0(r-v)} \rho_p(v,r) dv ds dr;$$

$$\Phi_{13}(c_0) = \int_0^\infty \int_0^r \int_0^s e^{c_0(r-v)} l_{1p}(r,v,c_0) dv ds dr + \int_0^\infty \int_0^\infty \int_0^{r\wedge s} e^{c_0(r+s-2v)} l_{p1}(r,s,c_0) dv ds dr;$$

$$\Phi_{14}(c_0) = \int_0^\infty \int_0^r e^{c_0(r-s)} l_{2p}(r,s,c_0) ds dr + \int_0^\infty \int_0^r e^{c_0(r-s)} l_{2p}(s,r,c_0) ds dr;$$

$$\Phi_{15}(c_0) = \int_0^\infty \int_0^r e^{c_0(r-s)} l_{3p}(r,s,c_0) ds dr + \int_0^\infty \int_0^r e^{c_0(r-s)} l_{3p}(s,r,c_0) ds dr;$$

$$\begin{aligned} \Phi_{22}(c_0) = & \int_0^\infty \int_0^\infty \int_0^{r\wedge s} e^{c_0(r+s-2v)} \rho_p(r,s) dv ds dr \\ & + \int_0^\infty \int_0^\infty \int_0^r \int_0^s e^{c_0(r-v)} e^{c_0(s-q)} \rho_p(r,q) \rho_p(s,v) dq dv ds dr; \end{aligned}$$

$$\begin{aligned} \Phi_{23}(c_0) = & \int_0^\infty \int_0^\infty \int_0^\infty \int_0^{s\wedge v} e^{c_0(s+v-2q)} \rho_p(r,s) l_{1p}(r,v,c_0) dq dv ds dr \\ & + \int_0^\infty \int_0^\infty \int_0^r \int_0^s e^{c_0(r-v)} e^{c_0(s-q)} \rho_p(r,q) l_{1p}(s,v,c_0) dq dv ds dr; \end{aligned}$$

$$\begin{aligned} \Phi_{24}(c_0) = & \int_0^\infty \int_0^r \int_0^\infty e^{c_0(r-s)} \rho_p(r,v) l_{2p}(v,s,c_0) dv ds dr \\ & + \int_0^\infty \int_0^r \int_0^\infty e^{c_0(r-s)} \rho_p(r,v) l_2(s,v,c_0) dv ds dr; \end{aligned}$$

$$\begin{aligned} \Phi_{25}(c_0) = & \int_0^\infty \int_0^r \int_0^\infty e^{c_0(r-s)} \rho_p(r,v) l_{3p}(v,s,c_0) dv ds dr \\ & + \int_0^\infty \int_0^r \int_0^\infty e^{c_0(r-s)} \rho_p(r,v) l_{3p}(s,v,c_0) dv ds dr; \end{aligned}$$

$$\Phi_{33}(c_0) = \int_0^1\int_0^1\int_0^{r\wedge s} e^{c_0(r+s-2v)} I_{1p}(r;s;c_0)\,dv\,ds\,dr + \int_0^1\int_0^1\int_0^r\int_0^s e^{c_0(r-v)}e^{c_0(s-q)} I_{1p}(r;q;c_0)\,I_{1p}(s;v;c_0)\,dq\,dv\,ds\,dr;$$

$$\Phi_{34}(c_0) = \int_0^1\int_0^r e^{c_0(r-s)} I_{2p}(r;s;c_0)\,ds\,dr + \int_0^1\int_0^r e^{c_0(r-s)} I_{3p}(r;s;c_0)\,ds\,dr;$$

$$\Phi_{35}(c_0) = \int_0^1\int_0^r e^{c_0(r-s)} I_{3p}(r;s;c_0)\,ds\,dr + \int_0^1\int_0^r e^{c_0(r-s)} I_{3p}(s;r;c_0)\,ds\,dr;$$

$$\Phi_{44}(c_0) = \left(\operatorname{vec}A_p(c_0)^{-1}\right)' \operatorname{vec} I_{4p}(c_0) + \operatorname{tr}\left(A_p(c_0)^{-1}B_p(c_0)'A_p(c_0)^{-1}B_p(c_0)\right);$$

$$\Phi_{45}(c_0) = \operatorname{tr}\left(A_p(c_0)^{-1}B_p(c_0)A_p(c_0)^{-1}B_p(c_0)'\right) + \operatorname{tr}\left(A_p(c_0)^{-1}B_p(c_0)A_p(c_0)^{-1}B_p(c_0)\right);$$

$$\Phi_{55}(c_0) = \operatorname{tr}\left(A_p(c_0)^{-1}B_p(c_0)A_p(c_0)^{-1}B_p(c_0)'\right) + \operatorname{tr}\left(A_p(c_0)^{-1}B_p(c_0)A_p(c_0)^{-1}B_p(c_0)\right).$$

Proof

The proof uses Theorem 8, and we sketch the proof here. First, a direct calculation shows that $EQ_{iT} = 0$. Let $\Phi_{nT}(c_0) = EQ_{iT}Q_{iT}'$. Notice that Q_{iT} are iid $(0;\Phi_{nT}(c_0))$ across i. As $T \to \infty$,

$$Q_{iT} \Rightarrow Q_i;$$

where

$$\begin{aligned}
Q_i &= (Q_{1i};Q_{2i};Q_{3i};Q_{4i};Q_{5i})' \\
Q_{1i} &= \int_0^1 J_{c_0;i}(r)\,dW_i(r) \\
Q_{2i} &= \int_0^1\int_0^1 J_{c_0;i}(r)\,\tilde{h}_p(r;s)\,dW_i(s)\,dr \\
Q_{3i} &= \int_0^1\int_0^1 I_1(r;s;c_0)\,dW_i(r)\,dW_i(s) - \nu(c_0) \\
Q_{4i} &= \int_0^1\int_0^1 I_2(r;s;c_0)\,dW_i(r)\,dW_i(s) - \operatorname{tr}\left(A_p(c_0)^{-1}B_p(c_0)\right) \\
Q_{5i} &= \int_0^1\int_0^1 I_3(r;s;c_0)\,dW_i(r)\,dW_i(s) - \operatorname{tr}\left(A_p(c_0)^{-1}B_p(c_0)\right).
\end{aligned}$$

Also, a direct calculation shows that as $T \to \infty$,

$$\Phi_{nT}(c_0) = EQ_{iT}Q_{iT}' \to EQ_iQ_i' = \Phi(c_0).$$

Let l be any (5×1) vector with $\|l\| = 1$. We consider two cases.

Case 1: If $l'\Phi(c_0)l > 0$.

To establish the desired result with a joint limit, we apply Theorem 7. Condition (i) holds because it is assumed that $l'\Phi(c_0)l > 0$. Conditions (ii) and (iv) hold because $\lim_n \frac{1}{n}\sum_{i=1}^n \beta_i = \beta > 0$. Finally condition (iii), viz.

$$(l'Q_{iT})^2 \text{ are uniformly integrable in } T;$$

holds because $(l'Q_{iT})^2 \Rightarrow (l'Q_i)^2$ as $T \to \infty$ by the continuous mapping theorem with $E(l'Q_{iT})^2 = l'\Phi_{nT}(c_0)l \to l'\Phi(c_0)l = E(l'Q_i)^2$; and by applying Theorem 5.4 of Billingsley (1968).

Case 2: If $l'\Phi(c_0)l = 0$. Since $l'\Phi_{nT}(c_0)l \to l'\Phi(c_0)l = 0$;

$$E\left(\frac{1}{\sqrt{n}}\sum_{i=1}\Omega_i(l'Q_{iT})\right)^2 = \left(\frac{1}{n}\sum_{i=1}\Omega_i^2\, l'\Phi_{nT}(c_0)l\right) \to 0;$$

which leads to

$$\frac{1}{\sqrt{n}}\sum_{i=1}\Omega_i(l'Q_{iT}) \to_p 0:$$

Therefore, by the Cramér-Wold device, it follows that

$$\frac{1}{\sqrt{n}}\sum_{i=1}\Omega_i Q_{iT} \Rightarrow N\left(0; a^2\Phi(c_0)\right): \blacksquare$$

7.3 Appendix C: Proofs of Section 4

Proof of Lemma 2.

We show separately the following

$$\frac{1}{n}\sum_{i=1}(m_{1iT}(c) - \Omega_i m_1(c)) \to_p 0; \tag{46}$$

and

$$\frac{1}{n}\sum_{i=1}(m_{2iT}(c) - \Omega_i m_2(c)) \to_p 0; \tag{47}$$

uniformly in c:

First, by definition and the triangle inequality, we have

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}(m_{1iT}(c) - \Omega_i m_1(c))\right| \\
=& \left|\frac{1}{n}\sum_{i=1}\left\{\begin{array}{l}\left(\frac{1}{T}\sum_{t=1}^T \varepsilon_{it}y_{it-1} - \Lambda_i\right) + \left(-\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T \varepsilon_{it}y_{is-1}\bar{h}_{pT}(t;s) - \Omega_i\omega_1(c_0)\right) \\ -\Omega_i(\omega_{1T}(c) - \omega_1(c)) - (c - c_0)\left(\frac{1}{T^2}\sum_{t=1}^T\left(\tilde{y}_{i,-1\,t}\right)^2 - \Omega_i\omega_2(c_0)\right)\end{array}\right\}\right| \\
&+\left|\frac{1}{n}\sum_{i=1}\left(\hat{\Omega}_i - \Omega_i\right)\right||\omega_{1T}(c)| + \left|\frac{1}{n}\sum_{i=1}\left(\hat{\Lambda}_i - \Lambda_i\right)\right| \\
\le& \left|\frac{1}{n}\sum_{i=1}\left(\frac{1}{T}\sum_{t=1}\varepsilon_{it}y_{it-1} - \Lambda_i\right)\right| \\
&+\left|\frac{1}{n}\sum_{i=1}\left(-\frac{1}{T^2}\sum_{t=1}\sum_{s=1}\varepsilon_{it}y_{is-1}\bar{h}_{pT}(t;s) - \Omega_i\omega_1(c_0)\right)\right| \\
&+|c - c_0|\left|\frac{1}{n}\sum_{i=1}\left(\frac{1}{T^2}\sum_{t=1}\left(\tilde{y}_{i,-1\,t}\right)^2 - \Omega_i\omega_2(c_0)\right)\right| \\
&+\left(\frac{1}{n}\sum_{i=1}\Omega_i\right)|\omega_{1T}(c) - \omega_1(c)| + \left|\frac{1}{n}\sum_{i=1}\left(\hat{\Omega}_i - \Omega_i\right)\right||\omega_{1T}(c)| + \left|\frac{1}{n}\sum_{i=1}\left(\hat{\Lambda}_i - \Lambda_i\right)\right| \\
=& \; I + II + III + IV + V + VI; \text{ say.}
\end{aligned}$$

Notice that two terms I and II are independent of c; and by Lemma 9 of Moon and Phillips (1999b), I, II $\to_p 0$ as $(n,T\to\infty)$. Next, III $\to_p 0$ uniformly in c because $\left|\frac{1}{n}\sum_{i=1}^n\left(\frac{1}{T^2}\sum_{t=1}^T\tilde{y}_{i,t-1}^2-\theta_i\omega_2(c_0)\right)\right|$ in term III is independent of c and also by Lemma 9 of Moon and Phillips (1999b), it converges in probability to zero as $(n,T\to\infty)$; and $|c-c_0|$ is a continuous function on the compact parameter set C. Finally, since $|\omega_{1T}(c)-\omega_1(c)|\to 0$ uniformly in c (by pointwise convergence and continuity on the compact set) and $\frac{1}{n}\sum_{i=1}^n\theta_i$ converges, IV $\to 0$ uniformly in c. Also, since $\frac{1}{n}\sum_{i=1}^n\left(\hat{\theta}_i-\theta_i\right)$, $\frac{1}{n}\sum_{i=1}^n\left(\hat{\Lambda}_i-\Lambda_i\right)=o_p(1)$; and $\sup_{c\in C}\omega_{1T}(c)<K$ for some finite K, terms V and VI converges in probability to zero uniformly in c. Therefore, $\frac{1}{n}\sum_{i=1}^n\left(m_{1iT}(c)-\theta_i m_1(c)\right)\to_p 0$ uniformly in c as $(n,T\to\infty)$.

Next, to prove (47), we write by definition

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n m_{2iT}(c)\\
&=\frac{1}{n}\sum_{i=1}^n\left(\frac{1}{T}\sum_{t=1}^T\varepsilon_{it}y_{it-1}-\Lambda_i\right)-(c-c_0)\frac{1}{n}\sum_{i=1}^n\frac{1}{T^2}\sum_{t=1}^T y_{it-1}^2\\
&\quad-\frac{1}{n}\sum_{i=1}^n\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T\Delta_c g_{pt}\varepsilon_{it}\right)'A_{pT}(c)^{-1}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^T\Delta_c g_{pt}y_{it-1}\right)\right]\\
&\quad+(c-c_0)\frac{1}{n}\sum_{i=1}^n\left[\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^T\Delta_c g_{pt}y_{it-1}\right)'A_{pT}(c)^{-1}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^T\Delta_c g_{pt}y_{it-1}\right)\right]\\
&\quad-\frac{1}{n}\sum_{i=1}^n\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T\Delta_c g_{pt}\varepsilon_{it}\right)'A_{pT}(c)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T g_{pt-1}D_{pT}^{-1}\varepsilon_{it}\right)\right]\\
&\quad+(c-c_0)\frac{1}{n}\sum_{i=1}^n\left[\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^T\Delta_c g_{pt}y_{it-1}\right)'A_{pT}(c)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T g_{pt-1}D_{pT}^{-1}\varepsilon_{it}\right)\right]\\
&\quad+(c-c_0)\frac{1}{n}\sum_{i=1}^n\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T\Delta_c g_{pt}\varepsilon_{it}\right)'A_{pT}(c)^{-1}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^T g_{pt-1}D_{pT}^{-1}y_{it-1}\right)\right]\\
&\quad-(c-c_0)^2\frac{1}{n}\sum_{i=1}^n\left[\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^T\Delta_c g_{pt}y_{it-1}\right)'A_{pT}(c)^{-1}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^T g_{pt-1}D_{pT}^{-1}y_{it-1}\right)\right]\\
&\quad+\frac{1}{n}\sum_{i=1}^n\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T\Delta_c g_{pt}\varepsilon_{it}\right)'A_{pT}(c)^{-1}B_{pT}(c)A_{pT}(c)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T\Delta_c g_{pt}\varepsilon_{it}\right)\right]\\
&\quad-(c-c_0)\frac{1}{n}\sum_{i=1}^n\left[\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^T\Delta_c g_{pt}y_{it-1}\right)'A_{pT}(c)^{-1}B_{pT}(c)A_{pT}(c)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T\Delta_c g_{pt}\varepsilon_{it}\right)\right]\\
&\quad-(c-c_0)\frac{1}{n}\sum_{i=1}^n\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T\Delta_c g_{pt}\varepsilon_{it}\right)'A_{pT}(c)^{-1}B_{pT}(c)A_{pT}(c)^{-1}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^T\Delta_c g_{pt}y_{it-1}\right)\right]
\end{aligned}
$$

$$+ (c - c_0)^2 \frac{1}{n}\sum_{i=1} \left[\left(\frac{1}{T\sqrt{T}} \sum_{t=1} \Delta_c g_{pt} y_{it-1} \right)' A_{pT}(c)^{-1} B_{pT}(c) A_{pT}(c)^{-1} \left(\frac{1}{T\sqrt{T}} \sum_{t=1} \Delta_c g_{pt} y_{it-1} \right) \right]$$
$$- \left(\frac{1}{n}\sum_{i=1} \Omega_i \right) \lambda_T(c) - \frac{1}{n}\sum_{i=1} \left(\hat{\Omega}_i - \Omega_i \right) \lambda_T(c) - \frac{1}{n}\sum_{i=1} \left(\hat{\Lambda}_i - \Lambda_i \right).$$

Since each element in $\Delta_c g_{pt}$ and $g_{pt-1} D_{pT}^{-1}$ satisfies the conditions for $f(x; c)$ and $g(x; c)$ in Lemma 9, the desired result in (47) follows by Lemma 9, $-\frac{1}{n}\sum_{i=1}^n \left(\hat{\Omega}_i - \Omega_i \right), \frac{1}{n}\sum_{i=1}^n \left(\hat{\Lambda}_i - \Lambda_i \right) = o_p(1)$ and boundedness of $\lambda_T(c)$ over the parameter set C. $\square$

Proof of Lemma 3.

The proof is similar to that of Lemma 2 is omitted. $\square$

Proof of Lemma 4.

Here we give only a sketch of the proof. The details of the calculation are quite similar to the proof of Lemma 9(b) with a replacement of the standardizing factor $\frac{1}{n}$ by $\frac{1}{\sqrt{n}}$ and the proof of Theorem 14 of Moon and Phillips (1999b).

First, using the BN decomposition of ε_{it} in (29) and of y_{it} in (33), we write

$$\frac{1}{\sqrt{n}} \sum_{i=1} m_{1iT}(c_0) = \frac{1}{\sqrt{n}} \sum_{i=1} \Omega_i \left(Q_{1iT} - Q_{2iT} \right) + \frac{1}{\sqrt{n}} \sum_{i=1} R_{1iT} + o_p(1) \tag{48}$$

and

$$\frac{1}{\sqrt{n}} \sum_{i=1} m_{2iT}(c_0) = \frac{1}{\sqrt{n}} \sum_{i=1} \Omega_i \left(Q_{1iT} - Q_{3iT} - Q_{4iT} + Q_{5iT} \right) + \frac{1}{\sqrt{n}} \sum_{i=1} R_{2iT} + o_p(1);$$

where R_{1iT} and R_{2iT} are relevant remainder terms generated by the BN decompositions y_{it-1} and ε_{it}. The $o_p(1)$ terms above hold because it is assumed that

$$\frac{1}{\sqrt{n}} \sum_{i=1} \left(\hat{\Omega}_i - \Omega_i \right), \frac{1}{\sqrt{n}} \sum_{i=1} \left(\hat{\Lambda}_i - \Lambda_i \right) = o_p(1).$$

Using similar arguments to those in the proof of Theorem 14 of Moon and Phillips (1999b), it is possible to show that

$$\frac{1}{\sqrt{n}} \sum_{i=1} R_{1iT} = O_p\left(\frac{n}{T} \right) = o_p(1); \tag{49}$$

and by applying arguments similar to those in the proof of (42) and (43), it is also possible to show that

$$\frac{1}{\sqrt{n}} \sum_{i=1} R_{2iT} = O_p\left(\frac{n}{T} \right) = o_p(1).$$

Then, it follows that

$$\frac{1}{\sqrt{n}}\sum_{i=1}\begin{pmatrix} m_{1iT}(c_0) \\ m_{2iT}(c_0)\end{pmatrix} = J'\left(\frac{1}{\sqrt{n}}\sum_{i=1}\psi_i Q_{iT}\right)J + o_p(1).$$

Finally, applying Lemma 4 with $c_{n0} = c_0$ (i.e., $\kappa = 0$), we obtain the desired result. ■

Proof of Lemma 5.

Part (a).

By definition and by the Cauchy-Schwarz inequality,

$$\begin{aligned} &\sup_{c\in C:|c-c_0|\le\gamma_{nT}} |B_{nT}R_{1nT}(c;c_0)| \\ \le\ & 2\|B_{nT}M_{nT}(c_0)\|\left\|\hat{W}\right\| \sup_{c\in C:|c-c_0|\le\gamma_{nT}}\left\|\frac{1}{n}\sum_{i=1} r_{iT}(c;c_0)\right\|. \end{aligned}$$

By Lemma 4 and Assumption 7, $2\|B_{nT}M_{nT}(c_0)\|\left\|\hat{W}\right\| = O_p(1)$. Thus, to complete the proof, it is enough to show that $\sup_{c\in C:|c-c_0|\le\gamma_{nT}}\left\|\frac{1}{n}\sum_{i=1}^n r_{iT}(c;c_0)\right\| = o_p(1)$. Notice by definition and the triangle inequality that

$$\begin{aligned} &\sup_{c\in C:|c-c_0|\le\gamma_{nT}}\left\|\frac{1}{n}\sum_{i=1} r_{iT}(c;c_0)\right\| \\ \le\ & \sup_{c\in C:|c-c_0|\le\gamma_{nT}}\left\|\frac{1}{n}\sum_{i=1}(dm_{iT}(c) - dm_{iT}(c_0))\right\| \\ \le\ & \sup_{c\in C}\left\|\frac{1}{n}\sum_{i=1}(dm_{iT}(c) - \psi_i dm(c))\right\| + \left\|\frac{1}{n}\sum_{i=1}(dm_{iT}(c_0) - \psi_i dm(c_0))\right\| \\ & \left|\frac{1}{n}\sum_{i=1}\psi_i\right| \sup_{c\in C:|c-c_0|\le\gamma_{nT}}\|dm(c) - dm(c_0)\|. \end{aligned} \tag{50}$$

Then, the first and the second terms in (50) are $o_p(1)$ by Lemma 3 and the last term in (50) is also $o_p(1)$ because $dm(c)$ is continuous in c and $\frac{1}{n}\sum_{i=1}^n\psi_i$ has a finite limit. Therefore $\sup_{c\in C:|c-c_0|\le\gamma_{nT}}\left\|\frac{1}{n}\sum_{i=1}^n r_{iT}(c;c_0)\right\| = o_p(1)$, as required.

Part (b).

The proof of Part (b) is similar to that of Part (a) and is omitted. ■

Proof of Theorem 2.

The proof is similar to the proof of Theorem 1 of Andrews (1999). Define $\hat{\lambda}_{nT} = B_{nT}(\hat{c} - c_0)$. Then,

$$\begin{aligned} o_p(1) &\le B_{nT}^2(Z_{nT}(c_0) - Z_{nT}(\hat{c})) \\ &= -H_{nT}\hat{\lambda}_{nT}^2 + 2H_{nT}(B_{nT}S_{nT})\hat{\lambda}_{nT} \\ &\quad - \hat{\lambda}_{nT}B_{nT}R_{1nT}(\hat{c};c_0) - \hat{\lambda}_{nT}^2 R_{2nT}(\hat{c};c_0). \end{aligned}$$

From Lemmas 3 and 4 and Assumption 7, we have $H_{nT}, H_{nT}^{-1} = O_p(1)$ and positive with probability one and $B_{nT}S_{nT} = O_p(1)$. Also, by Lemma 5, $B_{nT}R_{1nT}(\hat{c};c_0) = o_p(1)$ and $R_{2nT}(\hat{c};c_0) = o_p(1)$. Then,

$$o_p(1) \le -|\hat{\lambda}_{nT}|^2 + 2O_p(1)|\hat{\lambda}_{nT}| + |\hat{\lambda}_{nT}|o_p(1) + |\hat{\lambda}_{nT}|^2 o_p(1),$$

which is rearranged as

$$|\hat{\lambda}_{nT}|^2 \le 2O_p(1)\,|\hat{\lambda}_{nT}| + o_p(1):$$

Then, the required result

$$\hat{\lambda}_{nT} = O_p(1)$$

follows by relation (7.4) in Andrews (1999), page 1377. ∎

Proof of Theorem 3.

To complete the proof, it is enough to show (a) $B_{nT}(\hat{c} - c_0) = B_{nT}(\hat{c}_q - c_0) + o_p(1)$ and (b) $B_{nT}(\hat{c}_q - c_0) = \hat{\lambda}_{nT} + o_p(1)$:

Part (a). Recall that $\frac{B_{nT}S_{nT}}{H_{nT}} = O_p(1)$ by Lemmas 3 and 4 and Assumption 7. Then, it follows by the definition of $B_{nT}(\hat{c}_q - c_0)$ that

$$\left(B_{nT}(\hat{c}_q - c_0) - \frac{B_{nT}S_{nT}}{H_{nT}}\right)^2 \le \left(\frac{B_{nT}S_{nT}}{H_{nT}}\right)^2 = O_p(1);$$

which leads to

$$B_{nT}(\hat{c}_q - c_0) = \frac{B_{nT}S_{nT}}{H_{nT}} + O_p(1) = O_p(1):$$

So, we find that $\hat{c}_q$ is also $B_{nT}\ (=\sqrt{n}) -$ consistent. Then, by definition, we have

$$\begin{aligned} o_p(1) &\le B_{nT}^2 Z_{nT}(\hat{c}_q) - B_{nT}^2 Z_{nT}(\hat{c}) \\ &= \left(B_{nT}(\hat{c}_q - c_0) - \frac{B_{nT}S_{nT}}{H_{nT}}\right)^2 - \left(B_{nT}(\hat{c} - c_0) - \frac{B_{nT}S_{nT}}{H_{nT}}\right)^2 + o_p(1) \\ &\le o_p(1); \end{aligned}$$

where the $o_p(1)$ in the second line holds because $B_{nT}(\hat{c}_q - c_0);\ B_{nT}(\hat{c} - c_0) = O_p(1)$: So,

$$\left|\left(B_{nT}(\hat{c}_q - c_0) - \frac{B_{nT}S_{nT}}{H_{nT}}\right)^2 - \left(B_{nT}(\hat{c} - c_0) - \frac{B_{nT}S_{nT}}{H_{nT}}\right)^2\right| = o_p(1): \qquad (51)$$

Now, for given $\delta > 0$; set $\varepsilon = \delta^2$: Then, since $B_{nT}(\hat{c}_q - c_0)$ achieves the minimum of the quadratic function $f(\lambda) = \left(\lambda - \frac{B_{nT}S_{nT}}{H_{nT}}\right)^2$ on the closed interval $\{\lambda : B_{nT}(\bar{c} - c_0) \le \lambda \le -B_{nT}c_0\}$; it follows that $|B_{nT}(\hat{c} - c_0) - B_{nT}(\hat{c}_q - c_0)| > \delta$ implies

$$\left|\left(B_{nT}(\hat{c}_q - c_0) - \frac{B_{nT}S_{nT}}{H_{nT}}\right)^2 - \left(B_{nT}(\hat{c} - c_0) - \frac{B_{nT}S_{nT}}{H_{nT}}\right)^2\right| > \varepsilon:$$

Therefore

$$\begin{aligned} &P\left\{|B_{nT}(\hat{c} - c_0) - B_{nT}(\hat{c}_q - c_0)| > \delta\right\} \\ \le\ & P\left\{\left|\left(B_{nT}(\hat{c}_q - c_0) - \frac{B_{nT}S_{nT}}{H_{nT}}\right)^2 - \left(B_{nT}(\hat{c} - c_0) - \frac{B_{nT}S_{nT}}{H_{nT}}\right)^2\right| > \varepsilon\right\} \\ \to\ & 0; \end{aligned}$$

where the last convergence holds by (51); and we have completed the proof of Part (a).

Part (b). First we consider the case $c_0 \in C_0 = \{0\}$: For any $\delta > 0$;

$$\begin{aligned} &P\left\{\left|B_{nT}(\hat{c} - c_0) - \hat{\lambda}_{nT}\right| > \delta\right\} \\ \le\ & P\left\{\frac{B_{nT}S_{nT}}{H_{nT}} < B_{nT}(\underline{c} - c_0)\right\} + P\left\{\frac{B_{nT}S_{nT}}{H_{nT}} > -B_{nT}c_0\right\}: \end{aligned}$$

Since $\frac{B_{nT}S_{nT}}{H_{nT}} = O_p(1)$; for given $\varepsilon > 0$; we can choose K and $(n_0; T_0)$ such that

$$P\left\{ \left| \frac{B_{nT}S_{nT}}{H_{nT}} \right| > K \right\} < \varepsilon \text{ for all } n \geq n_0 \text{ and } T \geq T_0:$$

Choose $n_1 = \max\left\{ \left(\frac{K}{c_0 - \underline{c}}\right)^2; \left(\frac{K}{c_0}\right)^2; n_0 \right\}$: Then, whenever $n \geq n_1$ and $T \geq T_0$;

$$\begin{aligned} & P\left\{ \frac{B_{nT}S_{nT}}{H_{nT}} < B_{nT}(\underline{c} - c_0) \right\} + P\left\{ \frac{B_{nT}S_{nT}}{H_{nT}} > -B_{nT}c_0 \right\} \\ \leq \; & 2P\left\{ \left| \frac{B_{nT}S_{nT}}{H_{nT}} \right| > K \right\} \leq 2\varepsilon; \end{aligned}$$

and therefore,

$$P\left\{ \left| B_{nT}(\hat{c} - c_0) - \hat{\lambda}_{nT} \right| > \delta \right\} \leq 2\varepsilon;$$

as required. ■

7.4 Appendix D: Proofs of Section 5

Proof of Lemma 6

Part (a).

Part (a) holds by Lemma 4 with $c_0 = 0$ and by considering the marginal limiting distribution distribution of $\sqrt{n}M_{1nT}(0)$. ■

Part (b).

The proof of Part (b) is similar to the proof of Lemma 4, and we give only a sketch of the proof. By definition and by Assumption 6,

$$\sqrt{n}dM_{1nT}(0) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{1}{T^2}\sum_{t=1}^{T}\left(\tilde{y}_{i,t-1}\right)^2 - \sigma_i\frac{1}{T^2}\sum_{t=1}^{T}\left(\frac{t-s-1}{T}\right)\hbar_{1T}(t;s)\right] + o_p(1);$$

because of Assumption 6. Using the BN-decomposition of ε_{it}; we can decompose

$$\begin{aligned} & \frac{1}{T^2}\sum_{t=1}^{T}\left(\tilde{y}_{i,t-1}\right)^2 - \sigma_i\frac{1}{T^2}\sum_{t=1}^{T}\left(\frac{t-s-1}{T}\right)\hbar_{1T}(t;s) \\ = \; & \sigma_i Q_{6iT} + R_{iT}; \end{aligned}$$

where $x_{it} = \sum_{s=1}^{t} u_{is}$ with $x_{i0} = 0$;

$$\begin{aligned} Q_{6iT} = \; & \frac{1}{T^2}\sum_{t=1}^{T} x_{it-1}^2 - \frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T} x_{it-1}x_{is-1}\hbar_{1T}(t;s) \\ & - \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\frac{(t \wedge s) - 1}{T}\right)\hbar_{1T}(t;s); \end{aligned}$$

and R_{iT} is the remainder term. The specific forms of R_{1iT} can be found in the proof of Lemma 9 in Moon and Phillips (1999b). Then, by modifying the proof of Lemma 9 in Moon and Phillips (1999b) with the results in Appendix B2, it is possible to show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_{1iT} = O_p\left(\sqrt{\frac{n}{T}}\right) = o_p(1);$$

since $\frac{n}{T} \to 0$. Also, it is not difficult to prove that $Var(Q_{6iT}) \to \frac{11}{6300}$ as $(n,T) \to \infty$ for all i. Therefore, Part (b) holds. ■

Part (c).

Notice that

$$\begin{aligned}
\sqrt{n}\left(d^2 M_{1nT}(0)\right) &= \left(-\frac{1}{n}\sum_{i=1}^{n}\hat{\Omega}_i\right) d^2\omega_{1T}(0) \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Omega}_i\right)\left(\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\left(\frac{t-s-1}{T}\right)^2 \hbar_{1T}(t,s)\right).
\end{aligned}$$

From

$$\sup_{1\le t\le T}\ \sup_{\frac{t-1}{T}\le r\le \frac{t}{T}}\left|\left(\frac{t}{T}\right)^k - r^k\right| = \frac{1}{T}O(1) \text{ for all finite } k,$$

we have

$$\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\left(\frac{t-s-1}{T}\right)^2 \hbar_{1T}(t,s) \to \int_0^1\int_0^r (r-s)^2\,\hbar(r,s)\,ds\,dr + \frac{1}{T}O(1).$$

Also, a direct calculation shows that

$$\int_0^1\int_0^r (r-s)^2\,\hbar(r,s)\,ds\,dr = 0.$$

Therefore, since it is assumed that $\frac{n}{T} \to 0$ and $\frac{1}{n}\sum_{i=1}^{n}\hat{\Omega}_i \to_p \Omega$,

$$\sqrt{n}\left(d^2 M_{1nT}(0)\right) \to_p 0,$$

which is required. ■

Part (d).

By definition,

$$\begin{aligned}
d^3 M_{1nT}(c) &= \left(-\frac{1}{n}\sum_{i=1}^{n}\hat{\Omega}_i\right) d^3\omega_{1T}(c) \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Omega}_i\right)\left(\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1} e^{c\left(\frac{t-s-1}{T}\right)}\left(\frac{t-s-1}{T}\right)^3 \hbar_{1T}(t,s)\right).
\end{aligned}$$

Notice that $d^3M_{1nT}(c)$ is continuous on the compact parameter set. Since

$$\begin{aligned}
&\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1} e^{c\left(\frac{t-s-1}{T}\right)}\left(\frac{t-s-1}{T}\right)^3 \hbar_{1T}(t,s) \\
&\to d^3M_1(c,0) = \int_0^1\int_0^r e^{c(r-s)}(r-s)^3\,\hbar(r,s)\,ds\,dr
\end{aligned}$$

and $\frac{1}{n}\sum_{i=1}^{n}\hat{\Omega}_i \to_p \Omega$,

$$d^3M_{1nT}(c) \to_p \Omega d^3M_1(c,0)$$

uniformly in $c \in C$, and we have the required result. ■

Before we prove Lemma 7, we introduce the following lemma which is helpful in deriving the asymptotics of $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}^2 - \varepsilon_{it}^2\right)$.

Lemma 11 Suppose that assumptions in Lemmas 6 and 7 hold. Then, as $(n,T \to \infty)$ with $\frac{n}{T} \to 0$,

$$\sqrt{nT}\left(\hat{\rho}^{++} - \rho_0\right) = O_p(1),$$

where $\hat{\rho}^{++}$ is defined in (16).

Proof of Lemma 11

By definition,

$$\begin{aligned}
&\sqrt{nT}\left(\hat{\rho}^{++} - \rho_0\right)\\
= &\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\left(\underset{\sim}{y}_{i,-1}\right)_t^2\right)^{-1}\\
&\times\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\underset{\sim}{\varepsilon}_{it}\left(\underset{\sim}{y}_{i,-1}\right)_t - \Lambda_i - \Omega_i\,\omega_{1T}(0)\right)\right) + o_p(1),
\end{aligned}$$

where the $o_p(1)$ order holds because $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\tilde{\Lambda}_i - \Lambda_i\right), \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\tilde{\Omega}_i - \Omega_i\right) = o_p(1)$, and

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\left(\underset{\sim}{y}_{i,-1}\right)_t^2 = O_p(1) > 0.$$

Using Lemma 9(a) and (c), it is possible to show that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\left(\underset{\sim}{y}_{i,-1}\right)_t^2 \to_p \Omega\,\omega_2(0) = \frac{\Omega}{15}, \tag{52}$$

as $(n,T \to \infty)$. Next, notice that as $(n,T \to \infty)$ with $\frac{n}{T} \to 0$,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\underset{\sim}{\varepsilon}_{it}\left(\underset{\sim}{y}_{i,-1}\right)_t - \Lambda_i - \Omega_i\,\omega_{1T}(0)\right)\\
= &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}y_{it-1} - \Lambda_i\right)\\
&- \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\varepsilon_{it}y_{is-1}\tilde{h}_{1T}(t,s) + \Omega_i\,\omega_{1T}(0)\right)\\
= &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Omega_i\left(Q_{1inT} - Q_{2inT}\right) + o_p(1),
\end{aligned} \tag{53}$$

where the last equality holds by (48) and (49) with $c_0 = 0$ and $p = 1$, and Q_{1inT} and Q_{2inT} are the same in (44). In view of the proof of Lemma 10, the following holds

$$\limsup_{n,T} E\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Omega_i\left(Q_{1inT} - Q_{2inT}\right)\right)^2 < \infty. \tag{54}$$

Therefore, from (52), (53), and (54) the desired result follows. $\blacksquare$

Proof of Lemma 7

Part (a).

By definition, we can write

$$
\begin{aligned}
M_{2nT}(0) &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}y_{it-1} - \Lambda_i\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}y_{it-1}\right) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{t-1}{T}\varepsilon_{it}\right) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}\right)^2\left(\frac{1}{T}\sum_{t=1}^{T}\frac{t-1}{T}\right) - \left(\frac{1}{n}\sum_{i=1}^{n}\rho_{i\lambda_T}(0)\right) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\rho}_i - \rho_{i\lambda_T}(0)\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\Lambda}_i - \Lambda_i\right). \qquad (55)
\end{aligned}
$$

From the definitions of $\hat{\rho}_i$ and $\hat{\Lambda}_i$, the last two terms in (55) are

$$
\begin{aligned}
&-\left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\rho}_i - \rho_{i\lambda_T}(0)\right) + \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\Lambda}_i - \Lambda_i\right)\right) \\
&= \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}^2 - \sigma_{\varepsilon_i}^2\right)\right) + O_p\left(\frac{1}{T}\right). \qquad (56)
\end{aligned}
$$

Noticing that

$$
\left(\frac{1}{T}\sum_{t=1}^{T}y_{it-1}\varepsilon_{it}\right) - \Lambda_i = \frac{1}{2T}\left(y_{iT}^2 - y_{i0}^2\right) - \frac{1}{2T}\sum_{t=1}^{T}\left(\varepsilon_{it}^2 - \sigma_{\varepsilon_i}^2\right);
$$

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{t-1}{T}\varepsilon_{it} = \frac{T-1}{T}\frac{y_{iT}}{\sqrt{T}} - \frac{1}{T\sqrt{T}}\sum_{t=1}^{T}y_{it-1} + \frac{1}{T\sqrt{T}}y_{i0};
$$

and

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it} = \frac{y_{iT}}{\sqrt{T}} - \frac{y_{i0}}{\sqrt{T}};
$$

the other terms in (55) equal

$$
-\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_{it}^2 - \sigma_{\varepsilon_i}^2\right)\right) + O_p\left(\frac{1}{T}\right). \qquad (57)
$$

Putting (56) and (57) together, we have

$$
\begin{aligned}
&\sqrt{n}M_{2nT}(0) \\
&= -\frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_{it}^2 - \sigma_{\varepsilon_i}^2\right) + \frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}^2 - \sigma_{\varepsilon_i}^2\right) + o_p(1) \\
&= \frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}^2 - \varepsilon_{it}^2\right) + o_p(1).
\end{aligned}
$$

To show

$$
\frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}^2 - \varepsilon_{it}^2\right) = o_p(1);
$$

we write

$$\frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}^{2}-\varepsilon_{it}^{2}\right)=\sqrt{\frac{1}{T}}\left[\frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}^{2}-\varepsilon_{it}^{2}\right)\right].$$

By definition of $\hat{\varepsilon}_{it}$,

$$\begin{aligned}
&\frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}^{2}-\varepsilon_{it}^{2}\right),\\
=\;&\frac{1}{2\sqrt{nT}}\sum_{i=1}^{n}\left(\tilde{\varepsilon}_{i}'\tilde{\varepsilon}_{i}-\varepsilon_{i}'\varepsilon_{i}\right)-2\frac{1}{\sqrt{T}}\sqrt{nT}\left(\hat{\rho}^{++}-\rho_{0}\right)\frac{1}{nT}\sum_{i=1}^{n}\tilde{\varepsilon}_{i}'\tilde{y}_{i,-1}\\
&+\frac{1}{\sqrt{n}}\left(\sqrt{nT}\left(\hat{\rho}^{++}-\rho_{0}\right)\right)^{2}\left(\frac{1}{nT^{2}}\sum_{i=1}^{n}\tilde{y}_{i,-1}'\tilde{y}_{i,-1}\right)\\
=\;&\frac{1}{2\sqrt{nT}}\sum_{i=1}^{n}\left(\tilde{\varepsilon}_{i}'\tilde{\varepsilon}_{i}-\varepsilon_{i}'\varepsilon_{i}\right)+O_{p}\left(\frac{1}{\sqrt{T}}\right)+O_{p}\left(\frac{1}{\sqrt{n}}\right)\\
=\;&\frac{1}{2\sqrt{nT}}\sum_{i=1}^{n}\left(\tilde{\varepsilon}_{i}'\tilde{\varepsilon}_{i}-\varepsilon_{i}'\varepsilon_{i}\right)+o_{p}(1),
\end{aligned}$$

where the third line holds because $\frac{1}{nT}\sum_{i=1}^{n}\tilde{\varepsilon}_{i}'\tilde{y}_{i,-1}$, $\frac{1}{nT^{2}}\sum_{i=1}^{n}\tilde{y}_{i,-1}'\tilde{y}_{i,-1}=O_{p}(1)$ and $\sqrt{nT}\left(\hat{\rho}^{++}-\rho_{0}\right)=O_{p}(1)$ by Lemma 11.

Notice by definition that

$$\frac{1}{2\sqrt{nT}}\sum_{i=1}^{n}\left(\tilde{\varepsilon}_{i}'\tilde{\varepsilon}_{i}-\varepsilon_{i}'\varepsilon_{i}\right)=-\frac{1}{2}\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\varepsilon_{it}\varepsilon_{is}\hbar_{1T}(t,s)\right),$$

and using Lemma 9(d), it is possible to show that $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\varepsilon_{it}\varepsilon_{is}\hbar_{1T}(t,s)=O_{p}(1)$. So, since $\frac{n}{T}\to 0$,

$$\frac{1}{2\sqrt{nT}}\sum_{i=1}^{n}\left(\tilde{\varepsilon}_{i}'\tilde{\varepsilon}_{i}-\varepsilon_{i}'\varepsilon_{i}\right)=O_{p}\left(\sqrt{\frac{n}{T}}\right)=o_{p}(1);$$

and

$$\frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}^{2}-\varepsilon_{it}^{2}\right)=\frac{\sqrt{n}}{T}o_{p}(1)=o_{p}(1);$$

and we have desired result. ∎

Next, we sketch proofs for Parts (b) – (d). The details of the proofs for Part (b), (c), and (d) are similar to those of Part (b) of Lemma 6, Part (a) above, and Lemma 2, respectively, and we omit the details.

Part (b).

Taking the first derivative of $M_{2nT}(c)$ with respect to the parameter c, considering Assumption 6, and rearranging terms using the relations

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{t-1}{T}\varepsilon_{it}=\frac{y_{iT}}{\sqrt{T}}-\frac{1}{T}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}y_{it-1}+\frac{1}{T}\left(\frac{y_{i0}-y_{iT}}{\sqrt{T}}\right)\qquad(58)$$

and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it} = \frac{y_{iT}}{\sqrt{T}} - \frac{y_{i0}}{\sqrt{T}}; \tag{59}$$

it is possible to find that

$$\sqrt{n}d^2M_{2nT}(c) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\begin{array}{c} \left(-\frac{1}{T^2}\sum_{t=1}^{T}y_{it-1}^2 + \left(-\frac{1}{T}\sum_{t=1}^{T}\frac{t-1}{T}\right)\right) \\ +2\frac{y_{iT}}{\sqrt{T}}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\frac{t-1}{T}y_{it-1} - 2\left(-\frac{1}{T}\sum_{t=1}^{T}\frac{t-1}{T}\right)^2\right) \\ -\frac{1}{3}\left(\frac{y_{iT}}{\sqrt{T}}\right)^2 + \frac{1}{3}\left(-\right) \end{array}\right] + O_p\left(\sqrt{\frac{n}{T}}\right) + o_p(1).$$

Using the BN decomposition of y_{it-1} and the results in Appendix B2 with $c_0 = 0$; it is possible to show that

$$\begin{aligned} &\sqrt{n}dM_{2nT}(c) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[-\left(\begin{array}{c} \left(-\frac{1}{T^2}\sum_{t=1}^{T}x_{it-1}^2 + \frac{1}{T}\sum_{t=1}^{T}\frac{t-1}{T}\right) \\ +2\frac{x_{iT}}{\sqrt{T}}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\frac{t-1}{T}x_{it-1} - 2\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t-1}{T}\right)^2\right) \\ -\frac{1}{3}\left(\frac{x_{iT}}{\sqrt{T}}\right)^2 + \frac{1}{3} \end{array}\right)\right] + O_p\left(\sqrt{\frac{n}{T}}\right) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}Q_{7iT} + o_p(1); \end{aligned}$$

where $x_{it} = x_{it-1} + u_{it}$ with $x_{i0} = 0$. Then, direct calculations show that $EQ_{7iT} = 0$ and $Var(Q_{7iT}) \to \frac{1}{45}$. Therefore

$$\sqrt{n}dM_{2nT}(c) = O_p(1);$$

as required. ■

Part (c) and Part (d).

The proof of Part (c) is similar to that of Part (b). Taking the second order derivative of $M_{2nT}(c)$ with respect to the parameter c; considering Assumption 6, and rearranging terms using the relations of (58) and (59); it is possible to show that

$$\sqrt{n}d^2M_{2nT}(c) = O_p\left(\sqrt{\frac{n}{T}}\right) = o_p(1).$$

The proof of Part (d) is similar to the proof of Lemma 2. After taking the third order derivative of $M_{2nT}(c)$ with respect to c and using the results in Lemma 9, it is possible to show the required result. ■

Proof of Theorem 4

Define $\lambda_{nT} = n^{1/6}\hat{c}$. First, we consider the case where $\{|\lambda_{nT}| > 1\}$. By the definition of the GMM estimator, we have

$$\begin{aligned} o_p(1) &\le n\left(Z_{nT}(0) - Z_{nT}(\hat{c})\right) \\ &= -\sum_{k=1}^{2}\left(n^{(1-k/6)}A_{k;nT}\right)\lambda_{nT}^k - \sum_{k=3}^{6}\lambda_{nT}^k\left(n^{(1-k/6)}N_{k;nT}(\hat{c};0)\right). \end{aligned}$$

In view of (17) – (24) and from Assumption 7, λ_{nT} satisfies

$$o_p(1) \le -|\lambda_{nT}|^6 + |\lambda_{nT}|^5 o_p(1) + |\lambda_{nT}|^4 o_p(1) + 2O_p(1)|\lambda_{nT}|^3 + |\lambda_{nT}|^2 o_p(1) + |\lambda_{nT}| o_p(1). \tag{60}$$

Since, $|\hat{}_{nT}| > 1$;

$$\begin{aligned}&\text{The right hand side of (60)}\\ \le\ & -|\hat{}_{nT}|^6(1+o_p(1)) + 2O_p(1)|\hat{}_{nT}|^3.\end{aligned}$$

Then,

$$|\hat{}_{nT}|^6 \le 2O_p(1)|\hat{}_{nT}|^3 + o_p(1).$$

Following by relation (7.4) in Andrews (1999), page 1377, we can deduce that

$$|\hat{}_{nT}|^3 \le O_p(1) + o_p(1).$$

Therefore, when $\{|\hat{}_{nT}| > 1\}$;

$$|\hat{}_{nT}| \le O_p(1). \tag{61}$$

Finally, let the $O_p(1)$ random variable in (61) be ξ_{nT}. Then,

$$\begin{aligned}|\hat{}_{nT}| &= |\hat{}_{nT}|1\{|\hat{}_{nT}| \le 1\} + |\hat{}_{nT}|1\{|\hat{}_{nT}| > 1\}\\ &\le |\hat{}_{nT}|1\{|\hat{}_{nT}| \le 1\} + \xi_{nT}\\ &\le 1 + \xi_{nT} = O_p(1). \blacksquare\end{aligned}$$

Proof of Theorem 5

The proof of the theorem is similar to that of Theorem 3 and is omitted. $\blacksquare$

7.5 Appendix F: Numerical Validation of the Identification Condition of $m(c)$[6]

In this section we provide a numerical validation that the uniform limit of the moment conditions, $m(c) = (m_1(c); m_2(c))'$ has a root only at the true parameter $c = c_0$. We restrict the parameter set to $C = [-10; 0]$. The choice of the lower limit $\bar{c} = -10$ is made for computational convenience, and the results hold for all finite values of $\bar{c} < 0$. All the numerical analysis in this section is done with Mathematica and with Maple using Scientific Workplace Version 3.0.

7.5.1 When $g_{1t} = t$

The procedure we apply is to find all the roots of $m_2(c)$ and verify whether these roots are also the roots of $m_1(c)$. We first notice that for given c_0, the function $m_2(c)$ is simply the ratio of two polynomials - the denominator and the numerator of $m_2(c)$, say $m_{d2}(c)$ and $m_{n2}(c)$, respectively, are a fourth degree polynomial and a fifth degree polynomial in c, respectively.

Case A: When $c_0 \neq 0$

Step 1: Numerical Calculation of the roots of $m_2(c)$:

By a direct calculation, we find that the denominator of $m_2(c)$, $m_{d2}(c)$, equals to $4c_0^5\left(c^2 - 3c + 3\right)^2$ when $c_0 \neq 0$. Since $c^2 - 3c + 3 = \left(c - \frac{3}{2}\right)^2 + \frac{3}{4} > 0$, the denominator of $m_2(c)$ has no real roots for all $c_0 \neq 0$. Thus, if we concerned with the roots of $m_2(c)$, it suffices to consider only the numerator of $m_2(c)$, $m_{n2}(c)$. By definition of $m_2(c)$, we find that the true value $c = c_0$ is always a root of $m_{n2}(c)$. Also, by inspection, we find that $c = 0$ is always a root of $m_{n2}(c)$. Thus, we can write

$$m_{n2}(c) = c(c - c_0)\bar{m}_{n2}(c);$$

[6] We are in debt to John Owens for the numerical analysis in this section.

where $\tilde{m}_{n2}(c)$ is a third degree polynomial. Using Mathematica, we solve the third degree polynomial $\tilde{m}_{n2}(c)$ and find three roots of $\tilde{m}_{n2}(c)$ as a function of the true parameter c_0. For the numerical calculation we choose $\acute{c} = -10$, and so we assume that the parameter set $C = [-10, 0]$. The Figures A.1 and A.2 plot the graphs of these roots on C only when the roots are real numbers. As we see through the graphs, for $c_0 < 0$, the roots of $\tilde{m}_{n2}(c)$ are all positive, and so $\tilde{m}_{n2}(c)$ does not have a root in the parameter set C.

Step 2: Plug the bad root $c = 0$ of $m_2(c)$ to $m_1(c)$

We now investigate, for given $c_0 \in C \setminus \{0\}$, whether $m_1(c) = 0$ when $c = 0$. By matching the given true parameter c_0 with $m_1(0)$, we can define the function $m_1\_0(c_0)$ of c_0. Using Maple, we calculate

$$m_1\_0(c_0) = \frac{1}{4c^4}\left(\begin{array}{c} -c^3 + 48e^c - 8e^c c^2 - 8c^2 - 24 \\ +c^3 e^{2c} - 8e^{2c}c^2 + 24ce^{2c} - 24e^{2c} - 24c \end{array}\right),$$

and plot the graph of $m_1\_0(c_0)$. Figure A.3 plots $m_1\_0(c_0)$ on the range of $c_0 \in [-10, 0.4]$ and Figure A.4 plots the same function on the range of $c_0 \in [0.4, 0]$. Through these graphs, we can verify that $m_1\_0(c_0)$ is positive but very close to zero when the true value c_0 is close to zero.

Figure A.3 Graph of $m_1\_0(c_0)$ Figure A.4 Graph of $m_1\_0(c_0)$

To investigate further the behavior of $m_1\_0(c_0)$ around $c_0 = 0$, in Figure A.5 we plot the graphs of the first derivatives of numerator of $m_1\_0(c_0)$ on the range $c_0 \in [-0.05, 0]$.

Figure A.5. Graph of the first derivative of the Numerator of $m_1\_0(c_0)$

The graph shows that the first derivative of the numerator of $m_1\_0(c_0)$ is negative around zero, and so $m_1\_0(c_0)$ is strictly decreasing. Therefore, we conclude that $m_1\_0(c_0)$ is not zero for all $c_0 \in C_0$.

Case B: When $c_0 = 0$.

Using Maple, we calculate $m_2(c)$ when $c_0 = 0$, and plot the graph in Figures A.6 and A.7. From these figures, it is apparent that $m_2(c) = 0$ only when $c = c_0 = 0$.

Figure A.6 Graph of $m_2(c)$ when $c_0 = 0$ Figure A.7 Graph of $m_2(c)$ when $c_0 = 0$

7.5.2 When $g_{2t} = \left(t, t^2\right)$

Although the expressions involved in $m_2(c)$ in this case are far more complex, the analysis is simpler. Like the case of $g_{1t} = t$, we find that the denominator of $m_2(c)$ does not change sign over $C = [-10, 0]$, and so we focus on the numerator of $m_2(c)$. Similar to the case of $g_{1t} = t$, we numerically calculate the real roots of the numerator of $m_2(c)$ for $c_0 \in C = [-10, 0]$, and we find that there exists only one root in the range of c_0, which implies that $m_2(c) = 0$ only at the true c_0. Therefore, when $g_{2t} = \left(t, t^2\right)$, the limit of moment condition $m(c)$ identifies the true parameter c_0 in C.

References

[1] Andews, D.W.K. (1999) : Estimation When a Parameter Is on a Boundary, Econometrica, 67, 1341–1384.

[2] Billingsley, P. (1968): Convergence of Probability Measures, New York, Wiley.

[3] Bhattacharyya, A. (1946) : On Some Analogues of the Amount Information and Their Uses in Statistical Estimation, Sankhya, 8, 1–14, 201-218, 315-328.

[4] Bernard, A. and C. Jones (1996) : Productivity Across Industries and Countries: Time Series Theory and Evidence, Review of Economics and Statistics, 78, 135–146.

[5] Binder, M., C. Hsiao, and H. Pesaran (1999) : Likelihood Based Inference for Panel Vector Autoregressions: Testing for Unit Roots and Cointegration in Short Panels, Mimeo.

[6] Canjels, E. and M. Watson (1997) : Estimating Deterministic Trends in Presence of Serially Correlated Errors, Review of Economics and Statistics, 79, 184–200.

[7] Choi, I. (1999) : Unit Root Tests for Panel Data, forthcoming in Journal of International Money and Finance.

[8] Coakely, J., F. Kulasi, and R. Smith (1996) : Current Account Solvencyand the Feldstein-Horika Puzzle, Economic Journal, 106, 620–627.

[9] Davidson, J. (1994) : Stochastic Limit Theory, Oxford University Press.

[10] Hahn, J. (1998) : Asymptotically Unbiased Inference of Dynamic Panel Model with Fixed Effects When Both n and T are Large, Mimeo.

[11] Im, K., H. Pesaran, and Y. Shin (1996) : Testing for Unit Roots Heterogeneous Panels, Mimeo.

[12] Lancaster, T. (1998) : The Incidental Parameter Problem Since 1948, Mimeo.

[13] Levin, A. and C. Lin (1993) : Unit Root Tests in Panel Data: New Results, UCSD Working Paper.

[14] MacDonald, R. (1996) : Panel Unit Root Tests and Real Exchange Rates, Economics Letters, 50, 7–11.

[15] Maddala, G.S. and S. Wu (1999) : A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test, Oxford Bulletin of Economics and Statistics, 61, 631–652.

[16] Magnus, J. and H. Neudecker (1988) : Matrix Di¤erential Calculus, New York, Wiley.

[17] Moon, H.R. and P.C.B. Phillips (1998) : A Reinterpretation of the Feldstein-Horioka Regressions from a Nonstationary Veiwpoint, Mimeo.

[18] Moon, H.R. and P.C.B. Phillips (1999a) : Maximum Likelihood Estimation in Panels with Incidental Trends, Oxford Bulletin of Economics and Statistics, 61, 771–748.

[19] Moon, H.R. and P.C.B. Phillips (1999b) : Estimation of Autoregressive Roots near Unity using Panel Data, forthcoming in Econometric Theory.

[20] Moon, H.R. and P.C.B. Phillips (2000) : GMM Estimation of Autoregressive Roots Near Unity with Panel Data, Mimeo.

[21] Newey, W. and D. McFadden (1994) : Large Sample Estimation and Hypothesis Testing, in Engle R.F. and D. McFadden ed. Handbook of Econometrics Vol 4., North-Holland, Amsterdam, 2111-2245.

[22] Neyman, J. and E.L. Scott (1948) : Consistent Estimates Based on Partially Consistent Observations, Econometrica, 16, 1–32.

[23] Nickell, S. (1981) : Biases in Dynamic Models with Fixed E¤ects, Econometrica, 49, 1417–1426.

[24] Oh, K. Y. (1996) : Purchasing Power Parity and the Unit Root Tests Using Panel Data, Journal of International Money and Finance, 15, 405–418.

[25] Park, J. and P.C.B. Phillips (1988) : Statistical Infencence in Regressions with Integrated Processes, Part I, Econometric Theory, 4, 468–497.

[26] Pedroni, P. (1996) : Fully Modi…ed OLS for Heterogeneous Cointegrated Panels and the Case of Purchasing Power Parity, Indiana University Working Papers in Economics, No. 96-20.

[27] Pedroni, P. (1999) : Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors, Oxford Bulletin of Economics and Statistics, 61, 653–670.

[28] Pesaran, H. and R. Smith (1995) : Estimating Long-Run Relationships from Dynamic Heterogeneous Panels, Journal of Econometrics, 68, 79–113.

[29] Phillips, P. C. B. (1987) : Towards a Uni…ed Asymptotic Theory for Autoregression, Biometrica, 74, 535-547.

[30] Phillips, P. C. B. (1993) : Hyper-Consistent Estimation of a Unit Root in Time Series Regression. Cowles Foundation Discussion Paper #1040.

[31] Phillips, P. C. B. (1995) : Full Modi…ed Least Squares and Vector Autoregression. Econometrica, 63, 1023-1078.

[32] Phillips, P.C.B. and C. Lee (1996) : E¢ciency Gains from Quasi-Di¤erencing under Nonstationarity, in P.M. Robinson and M. Rosenblatt (eds.), Athens Conference of Applied Probability and Time Series: Volume II Time Series Analysis in Memory of E.J. Hannan, New York, Springer-Verlag.

[33] Phillips, P.C.B. and H.R. Moon (1999) : Linear Regression Limit Theory for Nonstationary Panel Data, Econometrica, 67, 1057–1111.

[34] Phillips, P.C.B. and V. Solo (1992) : Asymptotics for Linear Processes, Annals of Statistics, 20, 971–1001.

[35] Quah, D. (1994) : Exploiting Cross-Section Variations for Unit Root Inference in Dynamic Data, Economic Letters, 44, 9–19.

[36] Waterman, R. and B. Lindsay (1996) : Projected Score Methods for Approximating Conditional Scores, Biometrika, 83, 1–13.

[37] Waterman, R. and B. Lindsay (1998) : Projected Score Methods for Nuisance Parameters: Asymptotics and Neyman-Scott Problems, Mimeo.

[38] Wu, S. (1997) : Purching Power Parity Under the Currency Float: New Evidence from Panel Data Unit Root Tests, Mimeo.

[39] Wu, Y. (1996) : Are Real Exchange Rates Nonstationary? Evidence from A Panel Data Test, Journal of Money, Credit, and Banking, 28, 54–63.

Figure A.1. Graph of Roots of $\tilde{m}_{n2}(c)$

Figure A.2. Graph of Roots of $\tilde{m}_{n2}(c)$